Daily House Price Indexes: Volatility Dynamics and Longer-Run Predictions

by

Wenjing Wang

Department of Economics
Duke University

Date: ____________________________

Approved:

____
Tim Bollerslev, Co-Supervisor

____
Andrew Patton, Co-Supervisor

____
Patrick Bayer

____
Jia Li

____
George Tauchen

Dissertation submitted in partial fulfillment of the requirements for the degree of
Doctor of Philosophy in the Department of Economics
in the Graduate School of Duke University
2014

Abstract

Daily House Price Indexes: Volatility Dynamics and Longer-Run Predictions

by

Wenjing Wang

Department of Economics
Duke University

Date: ______________________
Approved:

____
Tim Bollerslev, Co-Supervisor

____
Andrew Patton, Co-Supervisor

____
Patrick Bayer

____
Jia Li

____
George Tauchen

An abstract of a dissertation submitted in partial fulfillment of the requirements for
the degree of Doctor of Philosophy in the Department of Economics
in the Graduate School of Duke University
2014

Abstract

This dissertation presents the construction procedure of “high-frequency” daily measure of changes in housing valuations, and analyzes its return dynamics, as well as investigates its relationship to capital markets. The dissertation consists of three chapters. The first chapter introduces the house price index methodologies and housing transaction data, and reviews the related literature. The second chapter shows the construction and modeling of daily house price indexes and highlights the informational advantage of the daily indexes. The final chapter provides detailed empirical and theoretical investigations of housing index return volatilities.

Chapter 2 discusses the relationship of the housing market with the other markets, such as consumer good market and financial markets. Different housing price indexes and their construction methodologies are introduced, with emphases on the repeat sales model and S&P/Case Shiller Home Price Index. A detailed description of the housing transaction data I use in the dissertation is also provided in this chapter.

Chapter 3 is co-authored with Professor Tim Bollerslev and Professor Andrew Patton. We construct daily house price indexes for ten major U.S. metropolitan areas. Our calculations are based on a comprehensive database of several million residential property transactions and a standard repeat-sales method that closely mimics the procedure used in the construction of the popular monthly Case-Shiller house price indexes. Our new daily house price indexes exhibit dynamic features similar to those of other daily asset prices, with mild autocorrelation and strong

conditional heteroskedasticity. The correlations across house price index returns are low at the daily frequency, but rise monotonically with the return horizon, and are commensurate with existing empirical evidence for existing monthly and quarterly house price series. Timely and accurate measures of house prices are important in a variety of applications, and are particularly valuable during times of turbulence, such as the recent housing crisis. To quantify the informational advantage of our daily index, we show that a relatively simple multivariate time series model for the daily house price index returns, explicitly allowing for commonalities across cities and GARCH effects, produces forecasts of monthly house price changes that are superior to various alternative forecast procedures based on lower frequency data.

Chapter 4 investigates the properties of housing index return volatilities. Similar to stock market volatility, housing volatilities are found to respond asymmetrically to negative and positive returns. A direct test of volatility on changes in loan-to-value ratio suggests that the observed volatility asymmetry does not stem from changes in degree of housing financial leverage, but could result from the risk premium carried by housing volatility, which is supported by a consumption-based asset pricing model with housing. Moreover, housing and stock volatilities are found to be positively correlated from a set of predictive regressions based on realized variances of housing and stock markets, in which higher (lower) volatility in one market will be followed by higher (lower) volatility in the other. Finally, housing and stock cross-sectional return dispersions are shown to contain useful information in predicting both within-market and cross-market realized volatilities.

Contents

List of Tables

List of Figures

Acknowledgements

I am indebted to my advisors, Tim Bollerslev and Andrew Patton, for their ideas, guidance and constant support. I am grateful to Professor Bollerslev and Professor Patton for engaging me into the research, which has become a part of this dissertation. Without their encouragement and incredible patience throughout my study and research, this dissertation would not have been possible.

I am also thankful to other members of my committee, George Tauchen, Jia Li, and Patrick Bayer for their valuable comments and instructions. I would like to thank Professor Bayer for valuable suggestions and providing me with the data. I would like to thank Professor Charles Becker for his career guidance, starting from my first year as a master's student at Duke, all the way through the job market process. I would like to thank all participants of Duke Financial Econometrics Lunch Group for helpful discussions. I would also like to express my sincere appreciation for the financial support from Department of Economics and Graduate School of Duke University.

Finally, I am grateful to my parents, Shaohua Wang and Xiaoyun Wang, for their unconditional love and support, and I dedicate this work to them.

1

Introduction

This dissertation is mainly motivated by the important effects of changes in housing valuations on the overall economy, which have been highlighted by the recent economic crisis. The level of housing prices is generally believed to be a reflection of consumer confidence, which is an important indicator of economic recovery.

Timely measures of housing price movements contain important information concerning the current state of economy, which should be of direct interest to policy makers, central banks, developers and lenders, as well as, potential buyers and sellers. Chapter 2, “Housing Market: Price Measures and Transaction Data,” introduces the existing measures of changes in housing valuations, which are called the house price indexes. They are only available at relatively low monthly or quarterly frequencies, compared to most other financial asset classes, which ignores the information in the within-month variations in housing prices and leads to an underestimate of housing market risk. Chapter 3, “Daily House Price Indexes: Construction, Modeling and Longer-Run Predictions” co-authored with my advisors Professor Tim Bollerslev and Professor Andrew Patton, directly addresses this problem by constructing a new set of daily house price indexes for ten major U.S. metropolitan areas, using a comprehensive database of more than a hundred millions housing transactions. Our daily indexes are based on standard repeat sales method that closely mimics the procedure used in the construction of the popular monthly Case-Shiller home price indexes. We find that the daily housing returns exhibit similar characteristics to other daily asset returns, with mild autocorrelation and strong conditional heteroskedasticity, which are well described by a relatively simple multivariate GARCH type model. The sample and model-implied correlations across house price index returns are very low at the daily frequency, but rise monotonically with the return aggregation horizon, and are all commensurate with existing empirical evidence for the existing monthly and quarterly house price series. We show the informational advantages of our new more finely sampled daily price series through forecasting performance comparison, in which daily housing returns, together with the daily return model, produce out-of-sample forecasts of monthly housing returns that are superior to various alternative forecast procedures based on lower frequency data.

The availability of daily house price indexes opens the possibility for many other applications. Chapter 4, "Housing Price Volatilities: Asymmetries and Linkage to Stock Price Volatilities," is based on the fact that daily housing data could afford a more accurate measure of housing volatilities, which is another important informational advantage of holding the finer sampled asset prices. Understanding the dynamics of housing volatility and its relation to financial market volatility is of great importance to portfolio design and risk management in the presence of housing or real estate securities. I find that similar to stock market volatility, housing volatility responses asymmetrically to negative and positive returns both in aggregate market and in seven out of the ten metropolitan areas. A direct test of this effect on changes in loan-to-value ratio suggests that the observed volatility asymmetry does not stem from changes in degree of underlying housing financial leverage, but could result from the risk premium carried by housing volatility. I provide a risk-based explanation of the aggregate housing volatility asymmetry through a stylized consumption-based asset pricing model with housing. The model also suggests that housing and stock volatilities are positively correlated, which I empirically examine using a set of predictive regressions based on realized variances of housing and stock markets. The empirical results suggest that volatilities of the two markets are strongly linked. In particular, higher (lower) volatility in one market will be followed by higher (lower) volatility in the other. Finally, housing and stock cross-sectional return dispersions are shown to contain useful information in predicting both within-market and cross-market realized volatilities.

2

Housing Market: Price Measures and Transaction Data

2.1 Introduction

The recent economic crisis, which arguably originated with the precipitous drop in housing prices that began in 2006, directly underscores the important effects of changes in housing valuation on the capital markets and the overall economy. For most of the U.S. households, residential home is their largest financial asset holdings in their portfolios, and thus changes in housing valuation influence their saving and spending decisions. The housing market not only affects the consumer goods market through wealth effect, it also influences the financial sector through mortgage market and mortgage-backed securities, as well as investors' portfolio management activities.

The total valuation of U.S. residential real estate market is about \$16.5 trillion in the year of 2010, according to the Federal Reserve Flow of Fund Accounts of the United States, comparing to about \$17.5 trillion as the total value of stock market, according to the Center for Research in Security Prices (CRSP). Case et al. (2005, 2011) examine the linkage between increases in housing wealth, financial wealth, and consumer spending. They find a statistically significant effect of housing wealth upon household consumption, which is larger than the effect of stock market wealth upon consumption.

2.1.1 Housing and financial markets

The housing market also influences the financial sector through mortgage market and investors' portfolio management activities. A number of studies discuss households' portfolio management in the presence of housing asset. Cocco (2005) studies the effect of investment in housing on the composition of portfolio, and finds that housing investment helps to explain the patterns of cross-sectional variation in the composition of wealth, as well as the level of stockholdings observed in portfolio composition data. Yao and Zhang (2005) compare the investors' portfolio compositions, when they own a house versus rent the housing services. They find that when indifferent between renting and owning a house, the investors who own a house reduce the equity proportion in their net worth, but hold a higher equity proportion in their liquid financial portfolio to reduce the financial and labor-income risks, which highlights the important role that housing asset plays in the investor's optimal dynamic portfolio decisions. Flavin and Yamashita (2002) examine the household's optimal portfolio problem in presence of owner-occupied housing asset under a mean-variance efficiency framework, and argue that the portfolio constraint imposed by the consumption demand of housing influences the household's optimal financial asset holdings. Flavin and Yamashita (2011) develop a model of optimal portfolio allocation, which accounts for the housing adjustment cost and housing price risks.

Housing prices are fundamental to many financial products, such as mortgage insurance, asset-backed securities and real estate investment trusts (REITs). Mortgage insurance is an insurance policy that compensates lenders for losses due to the default of a mortgage loan. Default rates are usually high when the housing prices continuously go down and leave the lenders or investors with negative equity. Bardhan et al. (2006) develop an option-based model for the pricing of mortgage insurance contracts in closed form, in which the sole state variable is the collateral value of the underlying asset. The dynamics of asset value would have significant impact on the pricing of the mortgage insurance contracts. Mortgage-backed security (MBS) is a sort of asset-backed security that is secured by a pool of mortgages. The mortgage default risk, which is one of the key valuation factor of MBS, is closely associated with the changes in housing prices. A real estate investment trust (REIT) is a company that owns, operates or invests in income-producing real estate. It is a liquid asset class traded on major exchanges. REITs are generally believed to share many similar characteristics with stocks, but are not as closely linked to the underlying real estate as expected, although the linkage is shown to be time-varying by Clayton and MacKinnon (2001). For example, Pavlov and Wachter (2010) show that a statistically significant relationship between REITs and real estate returns is only found in the office sector. This weak linkage might be attributable to the construction method of most real estate indexes. By examining the REITs, Cotter and Roll (2011) find that investment in real estate is far more risky than what might be inferred from the S&P/Case-Shiller indexes.

2.1.2 Housing price volatility

The effects of risk and valuation of housing market on the consumer goods market and financial markets underscore the importance of possessing an accurate measure of changes in housing valuations. The most common measure is the house price index, which tracks the average changes in housing prices in a particular geographic area. Various index construction methods are proposed and assessed in the housing literature, but those indexes are only measured at relatively low frequencies, either monthly or quarterly. The low frequency reporting makes the real-time monitoring of housing price movements impossible; in addition, the aggregation reduces the apparent volatility, thus resulting in chronic underestimation of housing risk. The diversification gains by including real estate and real estate derivatives (see, e.g., Webb et al., 1988; Hoesli et al., 2004) emphasize the importance of understanding the dynamics of housing price volatility and its relationship to the financial market volatility. The dynamics of housing return volatility itself have not received much attention in the literature. The aggregation or moving average procedure makes the housing prices very smooth, so housing is generally believed to be an investment asset with steady growth rate and low risk. The recent bubble, bust and gradual recovery in the U.S. housing market clearly reveal the significant volatility in housing prices. It would be particularly interesting to study how the dynamics of housing volatility interact with overall business conditions, such as the credit availability.

The volatility linkage between housing and financial markets is closely related to the extensive literature on volatility transmission. Over the past decade, most research has focused on the spillover effect among different geographic markets for the same asset class, for example equity, bonds, and foreign exchange (see, e.g., King and Wadhwani, 1990; Hamao et al., 1990; Baillie et al., 1993; Susmel and Engle, 1994). The recent turbulence in the U.S. housing market has brought attention to the volatility spillover among local housing markets and among real estate derivatives markets. Michayluk et al. (2006) study the behavior of return on synchronously priced indices of U.S. and U.K. securitized real estate markets, and show that there exist significant asymmetric effects on both the volatility and correlation dynamics between the two markets. Miao et al. (2011) find that volatility linkage tends to be stronger during the active phase than during the calm phase, and East region between New York, Boston and Washington, DC, show a considerable amount of volatility transmission, while housing markets in the Central and Mountain regions appear to be relatively independent. Hoesli and Reka (2011) examine volatility spillover effect between the stock and the securitized real estate markets for U.S., U.K., and Australia, among which U.S. is found to have the strongest effect. Zhu et al. (2012) focus on U.S. regional housing markets, and document that besides the geographic closeness, similarity in economic conditions is also an important source of cross-market dependencies. Much less attention has been paid to the volatility linkage between financial markets and real sectors, an exception being the stream of studies on the relationship between the equity and oil markets (Malik and Ewing, 2009; El Hedi Arouri et al., 2011). The lack of studies on the volatility linkage between the equity and housing markets is partly due to the lack of housing price measures that are at a comparable measurement frequency to equity data. The housing price measures are usually computed and published monthly or quarterly, while high frequency intraday stock market data has been commonly used in the stochastic volatility literature in recent years.

Houses are heterogeneous assets; every house is unique, in terms of location, attributes, and etc, which makes different houses respond differently to economic shocks. Homeowners also have different demographic characteristics. An increase in economic uncertainty increases the dispersion of some demographic characteristics, such as income due to heterogeneous abilities, and may also generate many corresponding housing activities, such as forced moves and foreclosures, both of which increase the cross-sectional dispersion of housing returns within a region, so the cross-sectional dispersions of housing returns should contain useful information of both housing market and overall macro economic conditions. The cross-sectional dispersion in the stock market, which relates to the idiosyncratic volatility, has been well studied in the finance literature. A study by Garcia et al. (2011) formally shows that the cross-sectional dispersion is a model-free measure of average idiosyncratic variance. The idiosyncratic volatility of the stock market is found to vary with economic conditions, and it also moves with stock market volatility together countercyclically (see, e.g. Campbell et al., 2001; Stivers, 2003; Connolly and Stivers, 2006). In terms of housing cross-sectional dispersion, Plazzi et al. (2008) document that the cross-sectional dispersion of commercial real estate returns fluctuates with macroeconomic variables that are closely related to business cycles, such as term and credit spread, inflation, and short-term interest rates. Van Nieuwerburgh and Weill (2010) use a spatial equilibrium model to study the house price dispersion across metropolitan areas, and argue that faced with an increase in the productivity dispersion across areas, households choose to reallocate from lower to higher productivity metropolitan areas, a choice that generates increases in the observed cross-sectional dispersions of both house prices and wages. By contrast, most measures of housing return dispersion, such as in Plazzi et al. (2008) and in Van Nieuwerburgh and Weill (2010), are computed across geographic areas. A more accurate measure of the housing risk faced by a typical homeowner should be based on the dispersion across the returns of specific houses. This measure will be constructed using a detailed housing transaction database and will be studied in this dissertation.

A good measure of changes in housing prices is essential to all the studies discussed above. Various house price indexes have been proposed in the real estate literature. Next section discusses the pros and cons of those indexes, with a focus on the popular S&P/Case-Shiller index that is based on a standard repeat sales model.

2.2 House price indexes

A house price index is a measure that tracks the average changes of house prices through time in a specific geographic area. It is of great importance to track the housing market movements and housing affordability. There are two major difficulties in measuring average house prices (see, e.g., Rappaport, 2007). One difficulty is the heterogeneity among houses. Every house is a unique asset, in terms of its location, characteristics, maintenance status, etc., which will be reflected in its price. Average house prices are to measure the price movements of house that has the average quality, with the assumption that average quality remains the same across time, although in practice, average quality has been increasing over time, because the newly built houses tend to have better quality or are more in line with current household's requirements than existing ones. However, the detailed house qualities are not always available or not directly observable, so when measuring average house prices, it is difficult to take the changing average qualities of houses into consideration. The other difficulty is sale infrequency. House is obviously not as a liquid asset class as stocks or bonds. For instance, the average sale interval is about 6 years in Los Angeles metropolitan area from our housing transaction database. The price of a house is not observed until actual transaction occurs and the houses sold at each point in time might not be a good representative sample of the overall housing stock.

Three main methodologies have been used in constructing house price index (Rappaport, 2007). The first one simply takes the median value of all transaction prices in the calculation period. The National Association of Realtors employ this methodology and publish median price of existing home sales monthly for both the national and four Census regions. The median price index has the advantages of calculation simplicity, but the median measure is highly sensitive to distributional changes and it does not control the changing quality of houses. The second methodology uses the hedonic technique, which prices the average quality house by explicitly pricing its mixed attributes. This method solves the changing quality problem, but requires much richer housing attributes data than typically available. The U.S. Census Bureau constructs its Constant Quality (Laspeyres) Price Index of New One-Family Houses Sold using the hedonic method and publishes quarterly. The third methodology is based on repeat sales model, which are used by Standard & Poor's and Federal Housing Finance Agency (FHFA) to publish their S&P/Case-Shiller index and FHFA House Price Index (HPI)[1]. S&P/Case-Shiller index and FHFA HPI are the most commonly used indexes by researchers, investors and media.

2.2.1 Hedonic pricing methods

The empirical hedonic pricing methods are derived from the early theoretical work by Rosen (1974) and Lancaster (1966). The hedonic models are based on the assumption that the individual perceives a house as a bundle of attributes, so the housing transactions prices reveal the marginal values of these attributes. The housing attributes include number of rooms, garage availability, housing location, neighborhood characteristics, etc. The hedonic models enable the empirically examination of values of each attribute through a straightforward regression approach, but they are also subject to some difficulties or limitations. One difficulty, for instance, would always be the concern of model misspecification, such as missing attribute variables or misspecification of the functional forms of the attributes. In addition, the overall quality of micro level housing data is not very high, which poses challenge for the empirically implementation of hedonic property value model. The Constant Quality (Laspeyres) Price Index of New One-Family Houses Sold provided by the U.S. Census Bureau is one of the well-known hedonic housing price indexes.

2.2.2 Repeat sales models

Repeat sales methodology is used to estimate house price changes by looking at repeat transactions of the same house, assuming that the quality of the same house remains the same over time unless there are records of significant renovations and reconstructions. The repeat sales model is first introduced by Bailey et al. (1963) and modified by Case and Shiller (1989). Shiller (1991) proposes arithmetic repeat sales model, which is currently used to construct S&P/Case-Shiller index.

[1] The FHFA House Price Index is earlier named as the Office of Federal Housing Enterprise Oversight (OFHEO) HPI.

The repeat sales models have two main advantages: it controls the heterogeneity in characteristics of houses and the estimation only requires data on transaction prices and sale dates for properties. The repeat sales models are also subject to several problems and limitations (see, e.g., Cho, 1996 for a detailed discussion). The first problem is the assumption that quality of identical property is unchanged over time. In practice, quality of most houses changes through aging, maintenance or reconstruction. Therefore, possible structural changes between two sales cause renovation bias. The second problems is that the repeat sales sample only consider the houses that have been sold at least twice during the sampling period. This small subset of houses might not be representative of all the houses sold or the entire housing stock at that point of time, which is referred as sample-selection bias. The third problem is that the time interval of estimation varies across different published indices. For instance, the FHFA HPI and S&P/Case-Shiller index are published monthly or quarterly. Calhoun et al. (1995) compare repeat sales index over various intervals, including annually, semiannually, quarterly as well as monthly and concludes that aggregation bias arises for all intervals greater than one month. One of the technical limitation is that the repeat sales index is a measure of relative price level rather than price level itself. The index captures the average price appreciation over time for a specific geographic area, but provides no direct guides of price levels across geographic areas. The other technical problem is that the repeat sale index is subject to continual revision. Index for each point of time are estimated simultaneously using all repeated transactions observed within the sample period. When new transactions arrive, new sale pairs will be formed, so all previous indices technically need to be re-estimated.

In order to solve the problems in repeat sales model, new generations of repeat sales models have been proposed in the literature. Cho (1996) provides a good survey for various new generations of the repeat sales models. The four main models are the intercept repeat sales model, the hybrid model, the distance-weighted repeat sales model and the autoregressive model. Shiller (1993) include the intercept to capture a temporal change in prices. The hybrid model (Case and Quigley, 1991; Quigley, 1995) combines hedonic model with repeat sales model and reduces the asymptotic forecasting standard errors. Goetzmann and Spiegel (1997) uses a distance weighting function defined in characteristic and geographical space to exploit the factor structure of error-covariance matrix of repeat sales model. Nagaraja et al. (2011) propose a autoregressive model using single and repeat sales. Their model decomposes transaction price into a fix time effect and a random ZIP code effect combined with an autoregressive component, then they construct house price index using the time effects. They demonstrate that their autoregressive model exhibits better prediction power than the standard repeat sales models. Although the new generation of repeat sales models have advantage over the conventional repeat sales model, the two most popular housing price indexes, S&P/Case-Shiller Home Price Index and Federal Housing Finance Agency (FHFA) HPI, both adopt the standard repeat sales model, since it is more computational efficient and less data quality demanding.

2.2.3 S&P/Case-Shiller Home Price Index

S&P/Case-Shiller index measures the price changes of single-family residential homes in 20 defined Metropolitan Statistical Areas (MSAs) and nationally. It is released monthly by Standard and Poor's. The index is a interval and value-weighted arithmetic repeat sales estimator proposed by Shiller (1991). The indexes that are constructed by modeling the price difference are called arithmetic indices, while those use log price differences are called geometric indexes[2]. The index by design is a value-weighted estimator. As Shiller (1991) argues, arithmetic rather than geometric average of prices and the value-weighted index provide a measure of the total

[2] The FHFA HPI is one of the most popular geometric house price indexes.

value of real estates in a portfolio. The construction of S&P/Case-Shiller index also includes a interval weighting procedure. As argued by Standard and Poor's, over longer sale intervals, there is increased variation in price changes for properties that are more likely to be caused by non-market factors, such as physical changes and idiosyncratic neighborhood effects, so sale pairs with longer transaction intervals are less likely to accurately represent the aggregate price movements of the housing market in a particular geographic area and should be assigned less weights in the index calculation.

Let there be $T + 1$ time periods where sale can occur from 0,1,...,T and t be the subscript for time. A sale pair is formed for a given house if it has been sold twice within the sample period. Sale pairs are formed to avoid overlapping in time periods. That is, for example, if a house has been sold three times, then there are two sale pairs for that house. One pair is based on the first and second transactions and the other pair is formed by the second and third transactions. For sale pair i, the model can be written as:

$$\beta_t P_{it} = \beta_s P_{is} + \sqrt{2}\sigma_\varepsilon \varepsilon_{it} + \sqrt{h}\sigma_\eta \eta_{it}, \quad T \geqslant t > s \geqslant 0 \tag{2.1}$$

where P_{it} is the sale price of house i at time t, β_t is the inverse of the index at time t and $h = t - s$ is the interval length between the two sales. The last two terms on the right-hand side together account for the heteroskedasticity of the errors in the sale pairs, where $\sqrt{2}\sigma_\varepsilon \varepsilon_{it}$ represents the mispricing error and $\sqrt{h}\sigma_\eta \eta_{it}$ captures the interval error. The mispricing error exists because of imperfect information between buyers and seller or random arrival of interested buyers. The interval error represents, as outlined previously, the drift over time of the price of an individual property away from the market trend. σ_ε and σ_h are standard deviations associated with those two types of errors. ε_{it} and η_{it} are independent and identically standard bivariate normal distributed.

S&P/Case-Shiller index is calculated using county record data. The sales pairing procedure collects repeat sales, arms-length transactions of single-family residential properties. Non-arms-length transactions and transactions where the property type designation is changed are excluded. Transactions with suspected data errors are also removed. Finally, repeat sales that occur within 6 months are excluded, since, as stated by Standard and Poor's, repeat sales within a short interval often indicate that transaction is non-arms-length, precedes or follows a substantial physical changes or is a fraudulent transaction.

The errors of the S&P/Case-Shiller model have heteroskedastic variance structure, so the indexes are estimated using a three-stage generalized least square type procedure (see Case and Shiller, 1987). The base period of S&P/Case-Shiller Index is January 2000, where the index value is set to equal to 100. All index points prior to the base period are estimated simultaneously. After the base period, the index points are estimated using a chain-weighting procedure, in which an index point is estimated conditional on all previously calculated indexes. The chain-weighting procedure is used to reduce continual revisions of published index points. Finally, the indices are computed in a rolling three-month window and it is implemented by including the transaction three times in three successive months. For instance, the December index is based on repeat sales data for October, November and December.

The metropolitan level S&P/Case-Shiller indexes can be used to construct the S&P/Case-Shiller Composite Indexes. Take the Composite 10 Index as an example. The Composite 10 Index is calculated by summing the products of metropolitan index levels and normalized weights, which are each area's share of the total aggregate value of housing stock in 10 areas in year 2000.

$$P_t^c = \sum_{i=1}^{10} w_i P_{i,t} \tag{2.2}$$

$$\text{where } w_i = V_{i,2000} / \sum_{i=1}^{10} V_{i,2000}$$

where $V_{i,2000} = S_{i,2000} \times P_{i,2000}$ and it is the product of the U.S. Census counts of units in area i ($S_{i,2000}$) and estimated average price of single-family residential properties in area i of year 2000 ($P_{i,2000}$). The S&P/Case-Shiller U.S. National Home Price Index, which tracks the value of single-family housing within the United States, is computed in a similar way, using the indexes for the nine U.S. Census divisions.

2.2.4 Federal Housing Finance Agency (FHFA) HPI

The Federal Housing Finance Agency (FHFA) House Price Indexes (HPI) are constructed using data provided by Fannie Mae and Freddie Mac. The indexes only use mortgage transactions on single-family properties. They are first published in 1995 by the Office of Federal Housing Enterprise Oversight (OFHEO), one of FHFA's predecessor agencies. This index is based on transactions involving conforming, conventional mortgages purchased or securitized by Fannie Mae or Freddie Mac. They are initially published quarterly, and beginning in March 2008, OFHEO started to publish monthly indexes for census divisions and the United States.

S&P/Case-Shiller Home Price Index and FHFA HPI are both based on repeat sales model, but there are three major differences in their data and construction methodologies. First, the data used in the two sets of indexes are different. The S&P/Case-Shiller indexes and FHFA's purchase-only series only use purchase prices in the index construction, while the all-transactions HPI also includes refinance appraisals. FHFA's data are from mortgages information provided by Fannie Mae and Freddie Mac, while the S&P/Case-Shiller indexes obtain information from county assessor and recorder offices. Second, the S&P/Case-Shiller indexes are value-weighted, arithmetic repeat sales estimator, while FHFA's indexes are equal-weighted geometric repeat sales estimator, so more expensive houses have larger impact on estimated index than cheaper houses in the S&P/Case-Shiller indexes. Finally, FHFA provides national, region, state and city level house price indexes, while Standard & Poor's does not have state level indexes.

2.3 Housing transaction data

The housing transaction data used in this dissertation is from DataQuick, a property information services. The DataQuick data contains a universe of detailed housing transactions, including new purchases and refinances, in the United States, involving more than a hundred million properties. I extract the housing transaction data from the largest ten U.S. metropolitan statistical areas, including Los Angeles, Boston, Chicago, Denver, Miami, Las Vegas, San Diego, San Francisco, New York and Washington, D.C. areas. Table A.1 shows the represented counties for each of the ten areas. For most of areas, the historical transaction records are from late 1990s to 2012, except for some large metropolitan areas, such as Los Angeles and Boston, where transactions are recorded from 1988. Properties are uniquely identified by property IDs in the DataQuick data. The property ID is crucial for our later daily index construction, since it enables us to identify sale pairs, which is the two successive transactions on the same house. The U.S. standard use codes contained in the DataQuick data are used to distinguish transactions of different types of properties, such as residential real estate and commercial real estate. Within the residential real estate class, properties can be further categorized into single-family houses, multi-family houses, apartments, and etc. Information about each transaction includes the transaction price and date, the names of the buyers and sellers, information about buyer's loan, as well as an indication whether transaction is arms-length. The loan information contains the amount of up to three loans that are associated with the property transaction. The arms-length transaction are those occur between two parties that each act on behalf of its own interest, while non arms-length transaction happens often among family members. Housing attributes, such as square footage, year of built, and number of bedrooms and bathrooms are recorded from the most recent tax assessment. Although there is limited information about all physical changes in houses between transactions, there are records of the year in which latest major house improvements are made.

Take Los Angeles-Long Beach-Santa Ana area (Los Angeles area[3]) as an example. This metropolitan area contains two counties, Los Angeles county and Orange county. The DataQuick data contains a total of 10,285,770 property transactions from January 1988 to October 2012 for this metropolitan area. 58% of them are single-family home transactions, among which about 43% are categorized as arm-length. Repeat sales models only use information of houses that have been sold at least twice in the sample period. In the data, about 35% of houses are sold only once, 33% are sold twice, 19% are sold three times and rest of them are sold more than three times. Following S&P/Case-Shiller's, sale pairs that have transaction time interval less or equal to six months are considered as high turnover frequency, and thus are excluded. Sale pairs are also removed if there are indications that major improvements have been made between the two transactions. As argued by S&P/Case-Shiller Methodology, the transaction prices from the house flippers might not be a good reflection of the prevailing market prices, and renovation changes the quality of houses, which violates the constant quality assumption that underlies the repeat sales model.

S&P/Case-Shiller's excludes transactions that contain suspect data errors where the values appear to be unrealistic and price anomalies relative to the statistical distribution of all price changes in the area. However, Standard and Poor's does not make public the detailed criterions used in detecting data errors and price anomalies. Following Caplin et al. (2008), we remove the transactions with sale price less

[3] For simplicity, the name of each area is abbreviated by the name of its largest county or city.

than \$5000 or greater than \$100,000,000 and sale pairs with annualized returns less than -50% to greater than 100%. Furthermore, we experimented with different definitions of price anormalies and then remove transactions with sale prices that are two standard deviations away from the mean of all sale prices within each month or are greater than 6 times the median of all prices within each month. The local deed offices are usually closed on weekends or federal holidays, so sale pairs with second transaction recorded at these dates are considered not accurate and are also removed. After all data cleaning procedure, for instance, there are a total of 877,885 qualified sale pairs for Los Angeles metro area from January 1988 to October 2012. The average turnover time is about 6 years and the average number of daily sale pairs is about 180.

3

Daily House Price Indexes: Construction, Modeling and Longer-Run Predictions

3.1 Introduction

For many U.S. households their primary residence represents their single largest financial asset holding: the Federal Reserve estimated the total value of the U.S. residential real estate market at \$16 trillion at the end of 2011, compared with \$18 trillion for the U.S. stock market (as estimated by the Center for Research in Security Prices). Consequently, changes in housing valuations importantly affect households' saving and spending decisions, and in turn the overall growth of the economy. A number of studies (e.g., Case et al., 2011) have also argued that the wealth effect of the housing market for aggregate consumption is significantly larger than that of the stock market. The recent economic crisis, which arguably originated with the precipitous drop in housing prices beginning in 2006, directly underscores this point. Despite all of of this, and in sharp contrast to most other financial asset classes, aggregate price indexes for residential real estate valuations are only available at relatively low monthly or quarterly frequencies.

Set against this background, we provide a new set of daily house price indexes for ten major U.S. metropolitan areas. To the best of our knowledge, this represents the first set of house price indexes at the daily frequency. Our construction is based on a comprehensive database of publicly recorded residential property transactions. We show that the dynamic dependencies in the new daily housing price series closely mimic those of other asset prices (see, e.g., Tsay, 2005, for a discussion of financial time series), and that these dynamic dependencies along with the cross-city correlations are well described by a standard multivariate GARCH type model. This relatively simple daily model in turn allows for the construction of improved longer-run monthly and quarterly housing price forecasts compared with forecasts based on existing monthly and/or quarterly indexes.

Our new daily house price indexes are based on the same "repeat-sales" methodology as the popular S&P/Case-Shiller monthly indexes (see Shiller, 1991), and the Office of Federal Housing Enterprise Oversight's quarterly indexes. However, the coarser monthly and quarterly frequency of reporting employed in the existing indexes ignores potentially important information in the daily records of housing transactions, and is likely to result in "aggregation biases" if the true index changes at a higher frequency than the measurement period. Aggregating the indexes to lower frequencies also reduces their volatility, thereby underestimating the true risk of the housing market.

More timely house prices are of direct interest to policy makers, central banks, developers and lenders alike. Also, even though actual housing decisions are made relatively infrequently, potential buyers and sellers could still benefit from more timely price indicators. The need for higher frequency daily indexing is perhaps most acute in periods when prices change rapidly, with high volatility, as observed during the recent financial crisis and its aftermath. To illustrate, Figure 3.1 shows our new daily house price index along with the oft-cited monthly S&P/Case-Shiller index for Los

Angeles from September 2008 through September 2010. The precipitous drop in the daily index over the first six months clearly leads the monthly index. Importantly, the daily index also shows the uptick in housing valuations that occurred around April 2009 some time in advance of the monthly index. Similarly, the more modest rebound that occurred in early 2010 is also first clearly manifest in the daily index.

Systematically analyzing the features of the dynamics of the new daily house price indexes for all of the ten metropolitan areas in our sample, we find that, in parallel to the daily returns on most other broadly defined asset classes, they exhibit only mild predictability in the mean, but strong evidence of volatility clustering. We show that the volatility clustering within and across the different house price indexes can be satisfactorily described by a multivariate GARCH model. The correlation between the daily returns on the city indexes is much lower than the correlation observed for the existing monthly return indexes. However, as we temporally aggregate the daily returns to monthly and quarterly frequencies, we find that the correlations increase to levels consistent with the ones observed for existing lower frequency indexes. Furthermore, we document that the new daily indexes do indeed result in improved forecasts, not solely in that they more quickly identifying turning points as suggested by Figure 3.1 for Los Angeles, but also more generally for longer forecast horizons and other sample periods. This holds true for the city-specific housing returns and a composite index, thus directly underscoring the informational advantages of the new daily index developed here vis-a-vis the existing monthly published indexes.

The rest of the paper is organized as follows. The next section provides a review of house price index construction and formally describes the S&P/Case-Shiller methodology. Section 3.3 describes the data and the construction of our new daily prices series. Section 3.4 briefly summarizes the dynamic and cross-sectional dependencies in the daily series, and presents our simple multivariate GARCH model designed to account for these dependencies. Section 3.5 demonstrates how the new daily series and our modeling thereof may be used in more accurately forecasting the corresponding longer-run returns. Section 3.6 concludes. A Supplemental Appendix contains additional details and empirical results.

3.2 House price index methodologies

The construction of house price indexes is plagued by two major difficulties. Firstly, houses are heterogeneous assets; each house is a unique asset, in terms of its location, characteristics, maintenance status, etc., all of which affect its price. House price indexes aim to measure the price movements of a hypothetical house of average quality, with the assumption that average quality remains the same across time. In reality, average quality has been increasing over time, because newly-built houses tend to be of higher quality and more in line with current households' requirements than older houses. Detailed house qualities are not always available or not directly observable, so when measuring house prices at an aggregate level, it is difficult to take the changing average qualities of houses into consideration. The second major difficulty is sale infrequency. For example, the average time interval between two successive transactions of the same property is about six years in Los Angeles, based on our data set described in Section 3.3 below. Related to that, the houses sold at each point in time may not be a representative sample of the overall housing stock.

Three main methodologies have been used to overcome the above-mentioned difficulties in the construction of reliable house price indexes (see, e.g., the surveys by Cho, 1996; Rappaport, 2007; Ghysels et al., 2013). The simplest approach relies on the median value of all transaction prices in a given period. The National Association of Realtors employ this methodology and publishes median prices of existing home sales monthly for both the national and four Census regions. The median price index has the obvious advantage of calculation simplicity, but it does not control for heterogeneity of the houses actually sold.

A second, more complicated, approach uses a hedonic technique, to price the "average quality" house by explicitly pricing its specific attributes. The U.S. Census Bureau constructs its Constant Quality (Laspeyres) Price Index of New One-Family Houses Sold using a hedonic method. Although this method does control for the heterogeneity of houses sold, it also requires much richer data than are typically available.

A third approach relies on repeat sales. This is the method used by both Standard & Poor's and the Office of Federal Housing Enterprise Oversight (OFHEC). The repeat sales model was originally introduced by Bailey, Muth, and Nourse (1963), and subsequently modified by Case and Shiller (1989). The specific model currently used to construct the S&P/Case-Shiller indexes was proposed by Shiller (1991) (see Clapp and Giaccotto, 1992; Meese and Wallace, 1997, for a comparison of the repeat-sales method with other approaches).

As the name suggests, the repeat sales method estimates price changes by looking at repeated transactions of the same house. This provides some control for the heterogeneity in the characteristics of houses, while only requiring data on transaction prices and dates. The basic models, however, are subject to some strong assumptions (see, e.g., the discussion in Cho, 1996; Rappaport, 2007). Firstly, it is assumed that the quality of a given house remains unchanged over time. In practice, of course, the quality of most houses changes through aging, maintenance or reconstruction. This in turn causes a so-called "renovation bias." Secondly, repeat sales indexes exploit information only from houses that have been sold at least twice during the sampling period. This subset of all houses may not be representative of the entire housing stock, possibly resulting in a "sample-selection bias." Finally, as noted above, all of the index construction methods are susceptible to "aggregation bias" if the true average house price fluctuates within the estimation window.

Our new daily home price indexes are designed to mimic the popular S&P/Case-

Shiller house price indexes for the "typical" prices of single-family residential real estate. They are based on a repeat sales method and the transaction dates and prices for all houses that sold at least twice during the sample period. If a given house sold more than twice, then only the non-overlapping sale pairs are used. For example, a house that sold three times generates included sale pairs from the first and second transaction, and the second and third transaction; the pair formed by the first and third transaction is not included.

Specifically, for a house j that sold at times s and t at prices $H_{j,s}$ and $H_{j,t}$, the repeat sales model postulates that,

$$\beta_t H_{j,t} = \beta_s H_{j,s} + \sqrt{2}\sigma_w w_{j,t} + \sqrt{(t-s)}\sigma_v v_{j,t}, \qquad 0 \leqslant s < t \leqslant T, \tag{3.1}$$

with the value of the house price index at time τ is defined by the inverse of β_τ. The last two terms on the right-hand side account for "errors" in the sale pairs, with $\sqrt{2}\sigma_w w_{j,t}$ representing the "mispricing error," and $\sqrt{(t-s)}\sigma_v v_{j,t}$ representing the "interval error." Mispricing errors are included to allow for imperfect information between buyers and sellers, potentially causing the actual sale price of a house to differ from its "true" value. The interval error represents a possible drift over time in the value of a given house away from the overall market trend, and is therefore scaled by the (square root of the) length of the time interval between the two transactions. The error terms $w_{j,t}$ and $v_{j,t}$ are assumed independent and identically standard normal distributed.

The model in (3.1) and the corresponding error structure naturally lend itself to estimation by a multi-stage generalized least square type procedure (for additional details, see Case and Shiller, 1987). The base period of the S&P/ Case-Shiller indexes is January 2000. All index values prior to the base period are estimated simultaneously. After the base period, the index values are estimated using a chain-weighting procedure that conditions on all previous values. This chain-weighting procedure is used to prevent revisions of previously published index values. Finally, the S&P/Case-Shiller indexes are smoothed by repeating a given transaction in three successive months, so that the index for a given month is based on sale pairs for that month and the preceding two months (see the Index Construction Section of S&P/Case-Shiller Home Price Index Methodology).

3.3 Daily house price indexes

We focus our analysis on the ten largest Metropolitan Statistical Areas (MSAs), as measured in the year 2000 (further details pertaining to the counties included in each of the ten MSAs are provided in Table A.1 of the Supplementary Appendix).

3.3.1 Data and data cleaning

The transaction data used in our daily index estimation is obtained from DataQuick, a property information company. This database contains detailed transactions of more than one hundred million properties in the United States. For most of the areas, the historical transaction records extends from the late 1990s to 2012, with some large metropolitan areas, such as Boston and New York, having transactions recorded as far back as 1987. Properties are uniquely identified by property IDs, which enable us to identify sale pairs. We rely U.S. Standard Use Codes contained in the DataQuick database to identify transactions of single-family residential homes.

Our data cleaning rules are based on the same filters used by S&P/Case-Shiller in the construction of their monthly indexes. In brief, we remove any transaction that are not "arms length," using a flag for such transactions available in the database. We also remove transactions with "unreasonably" low or high sale prices (below \$5000 or above \$100 million, and those generating an average annual return of below -50% or above 100%), as well as any sales pair with an interval of less than six months. Sale pairs are also excluded if there are indications that major improvements have been made between the two transactions, although such indications are not always present in the database. For the Los Angeles MSA, for example, this yields a total of 877,885 "clean" sale pairs, representing an average of 180 *daily* sale pairs over the estimation period. Additional details for all ten MSAs are provided in Table A.2 of the Supplementary Appendix.

3.3.2 Estimation of the daily index

The repeat-sales index estimation based on equation (3.1) is not computationally feasible at the daily frequency, as it involves the simultaneous estimation of several thousand parameters: the daily time spans for the ten MSAs range from 2837 for Washington D.C. to 4470 days for New York. To overcome this difficulty, we use an expanding-window estimation procedure: we begin by estimating daily index values for the final month in an initial start-up period, imposing the constraint that all of the earlier months in the period have only a single *monthly* index value. Restricting the daily values to be the same within each month for all but the last month drastically reduces the dimensionality of the estimation problem. We then expand the estimation period by one month, and obtain daily index values for the new "last" month. We continue this expanding estimation procedure through to the end of our sample period. (This estimation method results in an index that is "revision proof," in that earlier values of the index do not change when later data becomes available.) Finally, similar to the S&P/Case-Shiller methodology, we normalize all of the individual indexes to 100 based on their average values in the year 2000.

One benefit of the estimation procedure we adopt is that it is possible to formally test whether the "raw" daily price series actually exhibit significant intra-monthly variation. In particular, following the approach used by Calhoun et al. (1995) to test for "aggregation biases," we test the null hypothesis that the estimates of $\beta_{i,\tau}$ for MSA i are the same for all days τ within a given calendar month against the alternative that these estimates differ within the month. These tests strongly reject the null for all months and all ten metropolitan areas; further details concerning the actual F-tests are available upon request. We show below that this statistically significant intra-monthly variation also translates into economically meaningful variation and corresponding gains in forecast accuracy compared to the forecasts based on coarser monthly index values only.

3.3.3 Noise filtering

The raw daily house price indexes are subject to measurement errors, due to the relatively few transactions that are available on a given day. (The average number ranges from 49 for Las Vegas to 180 for Los Angeles.) To help alleviate this problem, it is useful to further clean the data in an effort to extract more accurate estimates of the the true latent daily price series. We rely on a standard Kalman filter-based approach to do so. Specifically, let $P_{i,t}$ denote the true latent index for MSA i at time t. We assume that the "raw" price indexes constructed in the previous section, $P^*_{i,t} = 1/\beta_{i,t}$, are related to the true latent price indexes by,

$$\log P^*_{i,t} = \log P_{i,t} + \eta_{i,t}, \tag{3.2}$$

where the $\eta_{i,t}$ measurement errors are assumed to be serially uncorrelated. For simplicity of the filter, we further assume that the true index follows a random walk with drift,

$$r_{i,t} \equiv \Delta \log P_{i,t} = \mu_i + u_{i,t}, \tag{3.3}$$

where $\eta_{i,t}$ and $u_{i,t}$ are mutually uncorrelated. It follows readily by substitution that,

$$r^*_{i,t} \equiv \Delta \log P^*_{i,t} = r_{i,t} + \eta_{i,t} - \eta_{i,t-1}. \tag{3.4}$$

Combining (3.3) and (3.4), this in turn implies an MA(1) error structure for the "raw" returns, with the value of the MA coefficient determined by the variances of

$\eta_{i,t}$ and $u_{i,t}$, σ_η^2 and σ_u^2. This simple MA(1) structure is consistent with the sample autocorrelations for the raw return series reported in Figure A.1 in the Supplementary Appendix.

Interpreting equations (3.3) and (3.4) as a simple state-space system, μ, σ_η^2 and σ_u^2 may easily be estimated by standard (quasi-)maximum likelihood methods. This also allows for the easy filtration of of the "true" daily returns, $r_{i,t}$, by a standard Kalman filter; see, e.g., Hamilton (1994). The Kalman filter implicitly assumes that $\eta_{i,t}$ and $u_{i,t}$ are *iid* normal. If the assumption of normality is violated, the filtered estimates are interpretable as best linear approximations. The Kalman filter parameter estimates reported in the Supplementary Appendix imply that the noise-to-signal (σ_η/σ_u) ratios for the daily index returns range from a low of 6.48 (Los Angeles) to a high of 15.18 (Boston), underscoring the importance of filtering out the noise.

The filtered estimates of the latent "true" daily price series for Los Angeles are depicted in Figure 3.2 (similar plots for all ten cities are available in Figure A.2 in the Supplementary Appendix). For comparison, we also include the raw daily prices and the monthly S&P/Case-Shiller index. Looking first at the top panel for the year 2000, the figure clearly illustrates how the filtered daily index mitigates the noise in the raw price series. At the same time, the filtered prices also point to discernable within month variation compared to the step-wise constant monthly S&P/Case-Shiller index.

The bottom panel of Figure 3.2 reveals a similar story for the full 1995-2012 sample period. The visual differences between the daily series and the monthly S&P/Case-Shiller index are obviously less glaring on this scale. Nonetheless, the considerable (excessive) variation in the raw daily prices coming from the noise is still evident. We will consequently refer to and treat the filtered series as *the* daily house price indexes in the sequel.

Before turning to our empirical analysis and modeling of the dynamic dependencies in the daily series, it is instructive to more formally contrast the information inherent in the daily indexes with the traditional monthly S&P/Case-Shiller index.

3.3.4 Comparisons with the monthly S&P/Case-Shiller index

Like the monthly S&P/Case-Shiller indexes, our daily house price indexes are based on all publicly available property transactions. However, the complicated non-linear transformations of the data used in the construction of the indexes prevent us from expressing the monthly indexes as explicit functions of the corresponding daily indexes. Instead, as a simple way to help gauge the relationship between the indexes, and the potential loss of information in going from the daily to the monthly frequency, we consider the linear projection of the monthly S&P/Case-Shiller returns for MSA i, denoted $r_{i,t}^{S\&P}$, on 60 lagged values of the corresponding daily index returns,

$$r_{i,t}^{S\&P} = \delta(L)r_{i,t} + \varepsilon_{i,t} \equiv \sum_{j=0}^{59} \delta_j L^j r_{i,t} + \varepsilon_{i,t}, \tag{3.5}$$

where $L^j r_{i,t}$ refers to the daily return on the j^{th} day before the last day of month t. (As discussed further below, all of the price series appear to be non-stationary. We consequently formulate the projection in terms of returns as opposed to the price levels.) The inclusion of 60 daily lags match the three-month smoothing window used in the construction of the monthly S&P/Case-Shiller indexes, discussed in Section 3.2. The true population coefficients in the linear $\delta(L)$ filter are, of course, unknown, however they are readily estimated by ordinary least squares (OLS).

The OLS estimates for $\delta_{j=0,\ldots,59}$ obtained from the single regression that pools the returns for all ten MSAs are reported in the top panel of Figure 3.3. Each of the individual coefficients are obviously subject to a fair amount of estimation error. At the same time, there is a clear pattern in the estimates for δ_j across lags, naturally suggesting the use of a polynomial approximation in j to help smooth out the estimation error. The solid line in the figure shows the resulting nonlinear least squares (NLS) estimates obtained from a simple quadratic approximation. The corresponding R^2s for the unrestricted OLS and the NLS fit ($\hat{\delta}_j = 0.1807 + 0.0101j - 0.0002j^2$) are 0.860 and 0.851, indicating only a slight deterioration in the accuracy of the fit by imposing a quadratic approximation to the lag coefficients. Moreover, even though the monthly S&P/Case-Shiller returns are not an exact linear function of the daily returns, the simple relationship dictated by $\delta(L)$ accounts for the majority of the monthly variation.

To further illuminate the features of the approximate linear filter linking the monthly returns to the daily returns, consider the gain and the phase of $\delta(L)$,

$$G(\omega) = \left[\sum_{j=0}^{59} \sum_{k=0}^{59} \delta_j \delta_k cos(|j-k|\omega) \right]^{1/2}, \quad \omega \in (0, \pi), \tag{3.6a}$$

$$\theta(\omega) = tan^{-1} \left(\frac{\sum_{j=0}^{59} \delta_j sin(j\omega)}{\sum_{j=0}^{59} \delta_j cos(j\omega)} \right), \quad \omega \in (0, \pi). \tag{3.6b}$$

Looking first at the gains in Figures 3.3b and 3.3c, the unrestricted OLS estimates and the polynomial NLS estimates give rise to similar conclusions. The filter effectively down-weights all of the high-frequency variation (corresponding to periods less than around 70 days), while keeping all of the low-frequency information (corresponding to periods in excess of 100 days). As such, potentially valuable information for forecasting changes in house prices is obviously lost in the monthly aggregate. Further along these lines, Figures 3.3d and 3.3e show the estimates of $\frac{\theta(\omega)}{\omega}$, or the number of days that the filter shifts the daily returns back in time across frequencies. Although the OLS and NLS estimates differ somewhat for the very highest frequencies, for the lower frequencies (periods in excess of 60 days) the filter systematically shifts the daily returns back in time by about 30 days. This corresponds roughly to one-half of the three month (60 business days) smoothing window used in the construction of the monthly S&P/Case-Shiller index.

In sum, the monthly S&P/Case-Shiller indexes essentially "kill" all of the within quarter variation inherent in the new daily indexes, while delaying all of the longer-run information by more than a month. We turn next to a more detailed analysis of the time series properties of the new daily indexes.

3.4 Time series modeling of daily housing returns

To facilitate the formulation of a multivariate model for all of the ten city indexes, we restrict our attention to the common sample period from June 2001 to September 2012. Excluding weekends and federal holidays, this yields 2,843 daily observations.

3.4.1 Summary statistics

Summary statistics for each of the ten daily series are reported in Table 3.1. Panel A gives the sample means and standard deviations for each of the index levels. Standard unit root tests clearly suggest that the price series are non-stationary, and as such the sample moments in Panel A need to be interpreted with care; further details concerning the unit root tests are available upon request. In the sequel, we therefore focus on the easier-to-interpret daily return series.

The daily sample mean returns reported in Panel B are generally positive, ranging a low of -0.006 (Las Vegas) to a high of 0.015 (Los Angeles and Washington D.C.). The standard deviation of the most volatile daily returns 0.599 (Chicago) is double that of the least volatile returns 0.291 (New York). The first-order autocorrelations are fairly close to zero for all of the cities, but the Ljung-Box χ^2_{10} tests for up to tenth order serial correlation indicate significant longer-run dynamic dependencies in many of the series.

The corresponding results for the squared daily returns reported in Panel C indicate very strong dynamic dependencies. This is also immediately evident from the plot of the ten daily return series in Figure 3.4, which show a clear tendency for large returns in an absolute sense to be follow by other large absolute returns. This directly mirrors the ubiquitous volatility clustering widely documented in the literature for other daily speculative returns (e.g., Tsay, 2005). Further, consistent with the evidence for other financial asset classes, there is also a clear commonality in the volatility patterns across the ten series.

3.4.2 Modeling conditional mean dependencies

The summary statistics discussed above point to existence of some, albeit relatively mild, dynamic dependencies in the daily conditional means for most of the cities. Some of these dependencies may naturally arise from a common underlying dynamic factor that influences housing valuations nationally. In order to accommodate both city specific and national effects within a relatively simple linear structure, we postulate the following model for the conditional means of the daily returns,

$$E_{t-1}(r_{i,t}) = c_i + \rho_{i1} r_{i,t-1} + \rho_{i5} r_{i,t-5} + \rho_{im} r^m_{i,t-1} + b_{ic} r^m_{c,t-1}, \tag{3.7}$$

where $r^m_{i,t}$ refers to the (overlapping) "monthly" returns defined by the summation of the corresponding daily returns,

$$r^m_{i,t} = \sum_{j=0}^{19} r_{i,t-j}, \tag{3.8}$$

and the composite (national) return $r_{c,t}$ is defined as a weighted average of the individual city returns,

$$r_{c,t} = \sum_{i=1}^{10} w_i r_{i,t}, \tag{3.9}$$

with the weights identical to the ones used in the construction of the composite ten city monthly S&P/Case Shiller index, which are 0.212, 0.074, 0.089, 0.037, 0.050, 0.015, 0.055, 0.118, 0.272, and 0.078. The own fifth lag of the returns is included to account for any weekly calendar effects. The inclusion of the own monthly returns and the composite monthly returns provides a parsimonious way of accounting for longer-run city-specific and common national dynamic dependencies. This particular formulation is partly motivated by the Heterogeneous Autoregressive (HAR) model proposed by Corsi (2009) for modeling so-called realized volatilities, and we will refer to it as an HAR-X model for short. We estimate this model for the conditional mean simultaneously with the model for the conditional variance described in the next section via quasi-maximum likelihood.

The estimation results in Table 3.2 reveal that the ρ_1 and ρ_5 coefficients associated with the own lagged returns are mostly, though not uniformly, insignificant when judged by the robust standard errors reported in parentheses. Meanwhile, the b_c coefficients associated with the composite monthly return are significant for nine out of the ten cities. Still, the one-day return predictability implied by the model is fairly modest, with the average daily R^2 across the ten cities equal to 0.024, ranging from a low of 0.007 (Denver) to a high of 0.049 (San Francisco). This mirrors the low R^2s generally obtained from time series modeling of other daily financial returns.

The adequacy of the common specification for the conditional mean in equation (3.7) is broadly supported by the tests for up to tenth-order serial correlation in the residuals $\varepsilon_{i,t} \equiv r_{i,t} - E_{t-1}(r_{i,t})$ from the model reported in Panel C of Table 3.2. Only two of the tests are significant at the 5% level (San Francisco and Washington, D.C.) when judged by the standard χ^2_{10} distribution. At the same time, the tests for serial correlation in the squared residuals $\varepsilon^2_{i,t}$ from the model, given in the bottom two rows of Panel C, clearly indicate strong non-linear dependencies in the form of volatility clustering.

3.4.3 Modeling conditional variance and covariance dependencies

Numerous parametric specifications have been proposed in the literature to describe volatility clustering in asset returns. Again, in an effort to keep our modeling procedures simple and easy to implement, we rely on the popular GARCH(1,1) model (Bollerslev, 1986) for describing the dynamic dependencies in the conditional variances for all of the ten cities,

$$Var_{t-1}(r_{i,t}) \equiv h_{i,t} = \omega_i + \kappa_i \varepsilon_{i,t-1}^2 + \lambda_i h_{i,t-1}. \tag{3.10}$$

The results from estimating this model jointly with the the conditional mean model described in the previous section are reported in Panel B of Table 3.2 together with robust standard errors following Bollerslev and Wooldridge (1992) in parentheses.

The estimated GARCH parameters are all highly statistically significant and fairly similar across cities. Consistent with the results obtained for other daily financial return series, the estimates for the sum $\kappa + \lambda$ are all very close to unity (and just above for Chicago, at 1.002) indicative of a highly persistent, but eventually mean-reverting, time-varying volatility process.

Wald tests for up to tenth-order serial correlation in the resulting standardized residuals, $\varepsilon_{i,t}/h_{i,t}^{1/2}$, reported in Panel C, suggest that little predictability remains, with only one city (San Francisco) rejecting the null of no autocorrelation. The tests for serial correlation in the squared standardized residuals, $\varepsilon_{i,t}^2/h_{i,t}$, reject the null for four cities, perhaps indicative of some remaining predictability in volatility not captured by this relatively simple model. However for the majority of cities the specification in equation (3.10) appears to provide a satisfactory fit. The dramatic reduction in the values of the test statistics for the squared residuals compared to the values reported in the second row of Panel C is particularly noteworthy.

The univariate HAR-X-GARCH models defined by equations (3.7) and (3.10) indirectly incorporate commonalities in the cross-city returns through the composite monthly returns $r_{c,t}$ included in the conditional means. The univariate models do not, however, explain the aforementioned commonalities in the volatilities observed across cities and the corresponding dynamic dependencies in the conditional covariances of the returns.

The Constant Conditional Correlation (CCC) model proposed by Bollerslev (1990) provides a particularly convenient framework for jointly modeling the ten daily return series by postulating that the temporal variation in the conditional covariances are proportional to the products of the conditional standard deviations. Specifically, let $\mathbf{r_t} \equiv [r_{1,t}, ..., r_{10,t}]'$ and $D_t \equiv diag\left\{h_{1t}^{1/2}, ..., h_{10,t}^{1/2}\right\}$ denote the 10×1 vector of daily returns and 10×10 diagonal matrix with the GARCH conditional standard deviations along the diagonal, respectively. The GARCH-CCC model for the conditional covariance matrix of the returns may then be succinctly expressed as,

$$Var_{t-1}(\mathbf{r_t}) = D_t R D_t, \tag{3.11}$$

where R is a 10×10 matrix with ones along the diagonal and the conditional correlations in the off-diagonal elements. Importantly, the R matrix may be efficiently estimated by the sample correlations for the 10×1 vector of standardized HAR-X-GARCH residuals; i.e., the estimates of $D_t^{-1}\left[\mathbf{r_t} - E_{t-1}(\mathbf{r_t})\right]$. The resulting estimates are reported in Table A.5 in the Supplementary Appendix.

We also experimented with the estimation of the Dynamic Conditional Correlation (DCC) model of Engle (2002), resulting in only a very slight increase in the maximized value of the (quasi-) log-likelihood function. Hence, we conclude that the relatively simple multivariate HAR-X-GARCH-CCC model defined by equations (3.7), (3.10), and (3.11) provides a satisfactory fit to the joint dynamic dependencies in the conditional first and second order moments of the ten daily housing return series.

3.4.4 Temporal aggregation and housing return correlations

The estimated conditional correlations from the HAR-X-GARCH-CCC model for the daily index returns reported in the Supplementary Appendix average only 0.022. By contrast the unconditional correlations for the monthly S&P/Case Shiller indexes calculated over the same time period average 0.708, and range from 0.382 (Denver–Las Vegas) to 0.926 (Los Angeles–San Diego). The discrepancy between the two sets of numbers may appear to call into question the integrity of our new daily indexes and/or the time-series models for describing the dynamic dependencies therein, however conditional daily correlations and the unconditional monthly correlations are not directly comparable. In an effort to more directly compare the longer-run dependencies inherent in our new daily indexes with the traditional monthly S&P/Case Shiller indexes, we aggregate our daily return indexes to a monthly level by summing the daily returns within a month (20 days). The unconditional sample correlations for these new monthly returns are reported in the lower triangle of Panel B in Table 3.3. These numbers are obviously much closer, but generally still below the 0.708 average unconditional correlation for the published monthly S&P/Case Shiller indexes.

However, as previously noted, the monthly S&P/Case Shiller indexes are artificially "smoothed," by repeating each sale pair in the two months following the actual sale. As such, a more meaningful comparison of the longer-run correlations inherent in our new daily indexes with the correlations in the S&P/Case Shiller indexes is afforded by the unconditional quarterly (60 days) correlations reported in the upper triangle of Panel B in Table 3.3. There, we find an average correlation of 0.668, and a range of 0.317 (Denver–Las Vegas) to 0.906 (Los Angeles–San Diego), which are quite close to the corresponding numbers for the published S&P/Case Shiller indexes.

These comparisons, of course, say nothing about the validity of the HAR-X-

GARCH-CCC model for the daily returns, and the low daily *conditional* correlations estimated by that model. As a further model specification check, we therefore also consider the model-implied longer-run correlations, and study how these compare with the sample correlations for the actual longer-run aggregate returns.

The top number in each element of Panels A and B of Table 3.3 gives the median model-implied unconditional correlations for the daily, weekly, monthly, and quarterly return horizons, based on 500 simulated sample paths. The bottom number in each element is the corresponding sample correlations for the actual longer-run aggregated returns. Although the daily unconditional correlations in Panel A are all close to zero, the unconditional correlations implied by the model gradually increase with the return horizon, and almost all of the quarterly correlations are in excess of one-half. Importantly, the longer-run model-implied correlations are all in line with their unconditional sample analogues.

To further illuminate this feature, Figure 3.5 presents the median model-implied and sample correlations for return horizons ranging from one-day to a quarter, along with the corresponding simulated 95% confidence intervals implied by the model for the Los Angeles–New York city pair. The model provides a very good fit across all horizons, with the actual correlations well within the confidence bands. The corresponding plots for all of the 45 city pairs, presented in Figure A.3 in the Supplementary Appendix, tell a similar story.

Taken as whole these results clearly support the idea that the longer-run cross-city dependencies inherent in our new finer sample daily house price series are consistent with those in the published coarser monthly S&P/Case Shiller indexes. The results also confirm that the joint dynamic dependencies in the daily returns are well described by the relatively simple HAR-X-GARCH-CCC model, in turn suggesting that this model could possibly be used in the construction of improved house price forecasts over longer horizons.

3.5 Forecasting housing returns

One of the major potential benefits from higher frequency data is the possibility of constructing more accurate forecasts by using models that more quickly incorporate new information. The plot for Los Angeles discussed in the introduction alludes to this idea. In order to more rigorously ascertain the potential improvements afforded by the daily house price series and our modeling thereof, we consider a comparison of the forecasts from the daily HAR-X-GARCH-CCC model with different benchmark alternatives.

Specifically, consider the problem of forecasting the 20-day ("monthly") return on the house price index for MSA i,

$$r_{i,t}^{(m)} \equiv \sum_{j=0}^{19} r_{i,t-j} \tag{3.12}$$

for forecast horizons ranging from $h = 20$ days ahead to $h = 1$ day ahead. When $h = 20$ this corresponds to a simple one-step ahead forecast for one-month returns, but for $h < 20$ an optimal forecast will contain a mixture of observed data and a forecast for the return over the remaining part of the month. We will use the period June 2001 to June 2009 as our in-sample period, and the period July 2009 to September 2012 as our out-of-sample period, with all of the model parameters estimated once over the fixed in-sample period.

Our simplest benchmark forecast is based purely on end-of-month data, and is therefore *not* updated as the horizon shrinks. We will consider a simple AR(1) for these monthly returns,

$$r_{i,t}^{(m)} = \phi_0 + \phi_1 r_{i,t-20}^{(m)} + e_{i,t}. \tag{3.13}$$

As the forecast is not updated through the month, the forecast made at time $t - h$

is simply the AR(1) forecast made at time $t-20$,

$$\hat{r}_{i,t-h}^{Mthly} = \hat{\phi}_0 + \hat{\phi}_1 r_{i,t-20}^{(m)}. \tag{3.14}$$

Our second benchmark forecast is again purely based on monthly data, but now we allow the forecaster to update the forecast at time $t-h$, which may be in the middle of a month. We model the incorporation of observed data by allowing the forecaster to take a linear combination of the monthly return observed on day $t-h$ and the one-month-ahead forecast made on that day,

$$\hat{r}_{i,t-h}^{Interp} = \left(1 - \frac{h}{20}\right) r_{i,t-h}^{(m)} + \frac{h}{20}\left(\hat{\phi}_0 + \hat{\phi}_1 r_{i,t-h}^{(m)}\right). \tag{3.15}$$

Our third forecast fully exploits the daily return information, by using the actual returns from time $t-19$ to $t-h$ as the first component of the forecast, as these are part of the information set at time $t-h$, and then using a "direct projection" method to obtain a forecast for the remaining h-day return based on the one-month return available at time $t-h$. Specifically,

$$\hat{r}_{i,t-h}^{Direct} = \sum_{j=h}^{19} r_{i,t-j} + \hat{\beta}_0^{(h)} + \hat{\beta}_1^{(h)} r_{i,t-h}^{(m)}, \tag{3.16}$$

where $\beta_0^{(h)}$ and $\beta_1^{(h)}$ are estimated from the projection:

$$\sum_{j=0}^{h-1} r_{i,t-j} = \beta_0^{(h)} + \beta_1^{(h)} r_{i,t-h}^{(m)} + u_{i,t}. \tag{3.17}$$

Finally, we consider a forecast based on the HAR-X-GARCH-CCC model presented in the previous section. Like the third forecast, this forecast uses the actual returns from time $t-19$ to $t-h$ as the first component, and then iterates the expression for the conditional daily mean in equation (3.7) forward to get forecasts for the remaining h days,

$$\hat{r}^{HAR}_{i,t-h} = \sum_{j=h}^{19} r_{i,t-j} + \sum_{j=0}^{h-1} \hat{E}_{t-h}\left[r_{i,t-j}\right]. \tag{3.18}$$

Given the construction of the target variable, we expect the latter three forecasts ("Interp", "Direct", "HAR") to all beat the "Mthly" forecast for all horizons less than 20 days. If intra-monthly returns have dynamics that differ from those of monthly returns, then we expect the latter two forecasts to beat the "Interp" forecast. Finally, if the HAR-X-GARCH-CCC model presented in the previous section provides a better description of the true dynamics than a simple direct projection, then we would expect the fourth forecast to beat the third.

Figure 3.6 shows the resulting Root Mean Squared Errors (RMSEs) for the four forecasts as a function of the forecast horizon, when evaluated over the July 2009 to September 2012 out-of-sample period. The first striking, though not surprising, feature is that exploiting higher frequency (intra-monthly) data leads to smaller forecast errors than a forecast based purely on monthly data. All three of the forecasts that use intra-monthly information out-perform the model based solely on end-of-month data. The only exception to this is for Las Vegas at the $h = 20$ horizon, where the HAR model slightly under-performs the monthly model.

Another striking feature of Figure 3.6 is that the more accurate modeling of the daily dynamic dependencies afforded by the HAR-X-GARCH-CCC model results in lower RMSEs across *all* forecasts horizons for eight of the ten cities. For San Francisco and Las Vegas the direct projection forecasts perform essentially as well as the HAR forecasts, and for Denver and Los Angeles the improvement of the HAR forecast is small (but positive for all horizons). For some of the cities (Boston, Miami and Washington D.C., in particular) the improvements are especially dramatic over longer horizons.

The visual impression from Figure 3.6 is formally underscored by Diebold-Mariano tests, reported in Table 3.4. Not surprisingly, the HAR forecasts significantly outperform the monthly forecasts for horizons of 1, 5 and 10 days, for all ten cities and the composite index. At the one-month horizon, a tougher comparison for the model, the HAR forecasts are significantly better than the monthly model forecasts for four out of ten cities, as well as the composite index, and are never significantly beaten by the monthly model forecasts. Almost identical conclusions are drawn when comparing the HAR forecasts to the "interpolation" forecasts, supporting the conclusion that the availability of daily data clearly holds the promise of more accurate forecasts, particularly over shorter horizons, but also even at the monthly level.

The bottom row of each panel in Table 3.4 compares the HAR forecasts with those from a simple direct projection model. Such forecasts have often been found to perform well in comparison with "iterated" forecasts from more complicated dynamic models. By contrast, the Diebold-Mariano tests reported here suggest that the more complicated HAR forecasts generally perform better than the direct projection forecasts. For no city-horizon pair does the direct projection forecast lead to significantly lower out-of-sample forecast RMSE than the HAR forecasts, while for many city-horizon pairs the reverse is true. In particular, for Boston, Miami and Washington D.C., the HAR forecasts significantly beat the direct projection forecasts across all four horizons, and for the composite index this is true for all but the shortest horizon.

3.6 Conclusion

We present a set of new *daily* house price indexes for ten major U.S. Metropolitan Statistical Areas spanning the period from June 2001 to September 2012. The indexes are based on the repeat sales method of Shiller (1991), and use a comprehensive database of several million publicly recorded residential property transactions. We demonstrate that the dynamic dependencies in the new daily housing price series closely mimic those of other financial asset prices, and that the dynamics, along with the cross-city correlations, are well described by a standard multivariate GARCH-type model. We find that this simple daily model allows for the construction of improved daily, weekly, and monthly housing price forecasts compared to the forecasts based solely on monthly price indexes.

The new "high frequency" house price indexes developed here open the possibility for many other applications. Most directly, by providing more timely estimates of movements in the housing market, the daily series should be of immediate interest to policy makers and central banks. In a related context, the series may also prove useful in further studying the microstructure of the housing market. At a broader level, combining the daily house price series with other daily estimates of economic activity should afford better and more up-to-date insights into changes in the macro economy. Along these lines, the series also hold the promise for the construction of more accurate forecasts for other macro economic and financial time series. We leave all of these issues for future research.

Table 3.1: Daily summary statistics

	Los Angeles	Boston	Chicago	Denver	Miami	Las Vegas	San Diego	San Francisco	New York	D.C.
					Panel A: Level					
Mean	177.764	145.561	128.901	118.049	162.896	136.511	164.473	137.702	159.450	170.039
Std. dev.	41.121	13.381	21.631	4.605	48.351	48.568	34.058	27.169	25.877	34.830
					Panel B: Returns					
Mean	0.015	0.008	-0.002	0.003	0.006	-0.006	0.010	0.005	0.011	0.015
Std.dev.	0.347	0.351	0.599	0.303	0.428	0.370	0.387	0.509	0.291	0.502
AR(1)	-0.059	0.047	0.008	-0.018	-0.034	0.061	-0.005	-0.113	0.049	-0.018
LB(10)	67.877	21.935	24.362	16.838	17.742	59.549	15.065	269.509	13.335	24.977
					Panel C: Squared returns					
Mean	0.121	0.123	0.358	0.092	0.183	0.137	0.150	0.259	0.085	0.252
Std. dev.	0.200	0.260	1.269	0.242	0.336	0.369	0.270	0.616	0.170	0.607
AR(1)	0.113	0.102	0.075	0.021	0.107	0.071	0.037	0.042	0.042	0.132
LB(10)	182.307	109.914	102.316	33.414	445.189	85.348	50.715	179.632	53.109	106.434

Note: The table reports summary statistics for each of the ten MSAs for the June 2001 to September 2012 sample period, a total of 2,843 daily observations. AR(1) denotes the first order autocorrelation coefficient. LB(10) refers to the Ljung-Box portmanteau test for up to tenth order serial correlation. The 95% critical value for this test is 18.31.

Table 3.2: Daily HAR-X-GARCH models

$$r_{i,t} = c_i + \rho_{i,1} r_{i,t-1} + \rho_{i,5} r_{i,t-5} + \rho_{i,m} r^m_{i,t-1} + b_{i,c} r^m_{c,t-1} + \varepsilon_{i,t}, \qquad \varepsilon_{i,t}|\Omega_{t-1} \sim N(0, h_{i,t})$$
$$h_{i,t} = \omega_i + \kappa_i \varepsilon^2_{i,t-1} + \lambda_i h_{i,t-1}$$

	Los Angeles	Boston	Chicago	Denver	Miami	Las Vegas	San Diego	San Francisco	New York	D.C.
					Panel A: Mean					
c ($\times 10^{-2}$)	1.710	-0.302	0.094	-0.074	1.152	-0.111	0.240	-0.222	0.908	1.245
	(0.678)	(0.769)	(0.163)	(5.338)	(0.942)	(0.368)	(3.221)	(0.223)	(0.538)	(0.884)
ρ_1	-0.080	0.030	0.005	-0.015	-0.034	0.004	-0.037	-0.094	0.040	0.012
	(0.020)	(0.022)	(0.011)	(0.052)	(0.020)	(0.016)	(0.020)	(0.018)	(0.020)	(0.024)
ρ_5	0.054	0.009	-0.006	0.010	-0.006	0.006	-0.036	0.160	0.004	0.032
	(0.020)	(0.017)	(0.010)	(0.101)	(0.032)	(0.039)	(0.022)	(0.022)	(0.017)	(0.020)
ρ_m	-0.014	-0.014	-0.023	-0.011	-0.008	0.017	-0.013	-0.014	-0.029	-0.035
	(0.007)	(0.005)	(0.007)	(0.008)	(0.006)	(0.004)	(0.006)	(0.006)	(0.006)	(0.007)
b_c	0.059	0.039	0.049	0.020	0.060	0.035	0.060	0.056	0.054	0.084
	(0.009)	(0.007)	(0.008)	(0.018)	(0.008)	(0.007)	(0.010)	(0.009)	(0.006)	(0.010)
R^2	0.039	0.018	0.009	0.007	0.027	0.044	0.030	0.049	0.033	0.027

	Los Angeles	Boston	Chicago	Denver	Miami	Las Vegas	San Diego	San Francisco	New York	D.C.
					Panel B: Variance					
ω ($\times 10^{-2}$)	0.013	0.230	0.075	0.215	0.016	0.014	0.024	0.023	0.041	0.067
	(0.015)	(0.074)	(0.058)	(0.103)	(0.014)	(0.013)	(0.028)	(0.026)	(0.023)	(0.043)
κ	0.020	0.056	0.056	0.034	0.013	0.017	0.014	0.016	0.026	0.032
	(0.008)	(0.010)	(0.009)	(0.012)	(0.003)	(0.006)	(0.007)	(0.006)	(0.005)	(0.006)
λ	0.979	0.926	0.946	0.943	0.986	0.982	0.985	0.983	0.969	0.965
	(0.009)	(0.012)	(0.009)	(0.017)	(0.002)	(0.006)	(0.008)	(0.007)	(0.006)	(0.007)
$\kappa + \lambda$	0.999	0.982	1.002	0.977	0.999	0.999	0.999	0.999	0.995	0.998
					Panel C: Serial correlation tests					
$\varepsilon_{i,t}$	16.325	10.934	15.178	11.144	8.952	18.086	8.953	25.641	7.133	18.906
	(0.091)	(0.363)	(0.126)	(0.346)	(0.537)	(0.054)	(0.537)	(0.004)	(0.713)	(0.042)
$\varepsilon^2_{i,t}$	92.430	62.011	56.910	22.875	150.471	46.849	41.513	72.156	36.577	36.247
	(0.000)	(0.000)	(0.000)	(0.011)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
$\varepsilon_{i,t} h^{-1/2}_{i,t}$	11.003	11.878	15.071	14.344	6.576	20.148	7.677	18.762	6.386	12.855
	(0.357)	(0.293)	(0.130)	(0.158)	(0.765)	(0.028)	(0.660)	(0.043)	(0.782)	(0.232)
$\varepsilon^2_{i,t} h^{-1}_{i,t}$	12.511	24.289	24.616	25.424	9.426	4.946	16.156	40.312	8.650	11.998
	(0.252)	(0.007)	(0.006)	(0.005)	(0.492)	(0.895)	(0.095)	(0.000)	(0.566)	(0.285)

Note: Panel A and B report Quasi Maximum Likelihood Estimates (QMLE) of HAR-X-GARCH models with robust standard errors in parentheses. Panel C reports Wald test statistics for up to tenth order serial correlation in the (squared) residuals and standardized residuals, with corresponding p-values in parentheses.

Table 3.3: Unconditional return correlations for different return horizons

	Los Angeles	Boston	Chicago	Denver	Miami	Las Vegas	San Diego	San Francisco	New York	D.C.
				Panel A: Daily (lower triangle) and Weekly (upper triangle)						
Los Angeles	–	0.117 0.065	0.061 0.124	0.066 0.073	0.197 0.219	0.172 0.250	0.198 0.240	0.280 0.309	0.164 0.145	0.156 0.204
Boston	0.017 0.026	–	0.033 0.068	0.068 0.128	0.139 0.130	0.133 0.121	0.143 0.063	0.118 0.054	0.105 0.128	0.120 0.129
Chicago	0.002 0.019	−0.007 −0.001	–	0.025 0.108	0.077 0.149	0.058 0.064	0.049 0.042	0.084 0.148	0.102 0.115	0.068 0.089
Denver	0.001 −0.003	0.023 0.031	−0.002 −0.003	–	0.105 0.100	0.092 0.110	0.100 0.090	0.060 0.106	0.053 0.006	0.084 0.090
Miami	0.072 0.069	0.047 0.043	0.024 0.046	0.044 0.047	–	0.173 0.239	0.178 0.214	0.165 0.176	0.187 0.169	0.150 0.183
Las Vegas	0.060 0.077	0.051 0.049	0.015 0.032	0.038 0.027	0.053 0.054	–	0.165 0.209	0.147 0.162	0.123 0.060	0.142 0.173
San Diego	0.077 0.072	0.059 0.053	−0.006 0.022	0.045 0.042	0.056 0.060	0.058 0.065	–	0.171 0.263	0.148 0.169	0.137 0.127
San Francisco	0.183 0.235	0.037 0.038	0.037 0.065	0.006 −0.003	0.057 0.060	0.052 0.068	0.069 0.066	–	0.138 0.137	0.136 0.151
New York	0.032 0.041	0.011 0.000	0.047 0.061	−0.009 −0.002	0.065 0.063	0.010 −0.002	0.027 0.029	0.024 0.031	–	0.149 0.088
Washington, D.C.	0.047 0.045	0.038 0.034	0.017 0.024	0.032 0.041	0.041 0.038	0.049 0.034	0.033 0.027	0.038 0.038	0.044 0.043	–

	Los Angeles	Boston	Chicago	Denver	Miami	Las Vegas	San Diego	San Francisco	New York	D.C.
	Panel B: Monthly (lower triangle) and Quarterly (upper triangle)									
Los Angeles	–	0.634 0.621	0.530 0.602	0.463 0.506	0.730 0.852	0.600 0.837	0.731 0.906	0.724 0.834	0.759 0.747	0.733 0.856
Boston	0.382 0.348	–	0.451 0.655	0.400 0.559	0.616 0.507	0.533 0.522	0.624 0.673	0.594 0.623	0.643 0.735	0.627 0.688
Chicago	0.266 0.344	0.207 0.320	–	0.323 0.502	0.519 0.612	0.417 0.510	0.513 0.567	0.500 0.667	0.572 0.767	0.532 0.675
Denver	0.251 0.355	0.210 0.254	0.138 0.293	–	0.457 0.370	0.391 0.317	0.454 0.557	0.416 0.625	0.458 0.411	0.456 0.513
Miami	0.493 0.619	0.384 0.277	0.274 0.355	0.271 0.239	–	0.591 0.797	0.696 0.769	0.669 0.754	0.734 0.761	0.697 0.801
Las Vegas	0.395 0.633	0.328 0.322	0.210 0.233	0.229 0.201	0.404 0.547	–	0.589 0.782	0.558 0.657	0.599 0.659	0.582 0.708
San Diego	0.497 0.626	0.388 0.307	0.260 0.276	0.266 0.351	0.468 0.570	0.400 0.497	–	0.678 0.822	0.731 0.711	0.694 0.824
San Francisco	0.511 0.623	0.334 0.288	0.253 0.404	0.216 0.427	0.424 0.527	0.343 0.417	0.435 0.600	–	0.700 0.663	0.677 0.791
New York	0.505 0.478	0.384 0.415	0.318 0.427	0.247 0.149	0.499 0.496	0.383 0.354	0.480 0.430	0.431 0.394	–	0.738 0.761
Washington, D.C.	0.469 0.603	0.366 0.375	0.277 0.385	0.253 0.309	0.444 0.515	0.368 0.444	0.433 0.551	0.414 0.486	0.478 0.437	–

Note: Model-implied correlations are upper numbers and data-based correlations are in smaller font just below. Daily, weekly, monthly and quarterly horizons correspond to 1, 5, 20, 60 days respectively.

Table 3.4: Diebold-Mariano forecast comparison tests

	Composite	LA	Boston	Chicago	Denver	Miami	Las Vegas	San Diego	SF	New York	D.C.
					Panel A: One-day-ahead ($h = 1$)						
Mthly v.s. HAR	9.240	8.337	7.378	10.060	9.845	8.680	9.981	9.929	8.067	8.981	9.142
Interp v.s. HAR	8.707	10.171	7.623	6.242	9.249	11.415	8.569	10.786	7.865	8.609	10.293
Direct v.s. HAR	1.599	1.381	2.943	-0.176	1.224	2.785	0.126	-0.276	3.139	-0.012	2.173
					Panel B: One-week-ahead ($h = 5$)						
Mthly v.s. HAR	4.956	4.458	3.876	5.412	5.126	5.087	6.682	6.581	4.258	5.268	4.981
Interp v.s. HAR	4.071	2.964	4.856	5.466	6.724	5.882	4.501	4.761	5.349	4.588	5.304
Direct v.s. HAR	4.495	1.200	3.580	1.514	1.141	2.669	-0.298	0.768	-0.373	0.562	3.212
					Panel C: Two-weeks-ahead ($h = 10$)						
Mthly v.s. HAR	4.544	2.751	3.799	6.647	4.343	4.078	5.204	5.847	3.453	5.261	4.392
Interp v.s. HAR	4.372	1.478	3.617	4.586	4.042	3.333	2.489	3.598	2.954	2.973	3.798
Direct v.s. HAR	5.668	0.828	3.567	2.640	0.763	2.585	-0.214	1.342	-0.381	0.964	3.563
					Panel D: One-month-ahead ($h = 20$)						
Mthly v.s. HAR	–	–	–	–	–	–	–	–	–	–	–
Interp v.s. HAR	6.762	0.623	3.553	4.117	0.830	2.211	-0.511	1.777	0.941	1.909	4.268
Direct v.s. HAR	–	–	–	–	–	–	–	–	–	–	–

Note: The table reports the Diebold-Mariano test statistics for equal predictive accuracy against the alternative that the HAR forecast outperforms the other three forecasts, Mthly, Interp and Direct. The test statistics are asymptotically standard Normal under the null of equal predictive accuracy. The tests are based on the out-of-sample period from July 2009 to September 2012. The Mthly, Interp and Direct models are all identical when $h = 20$, so only one set of test statistics are reported in Panel D.

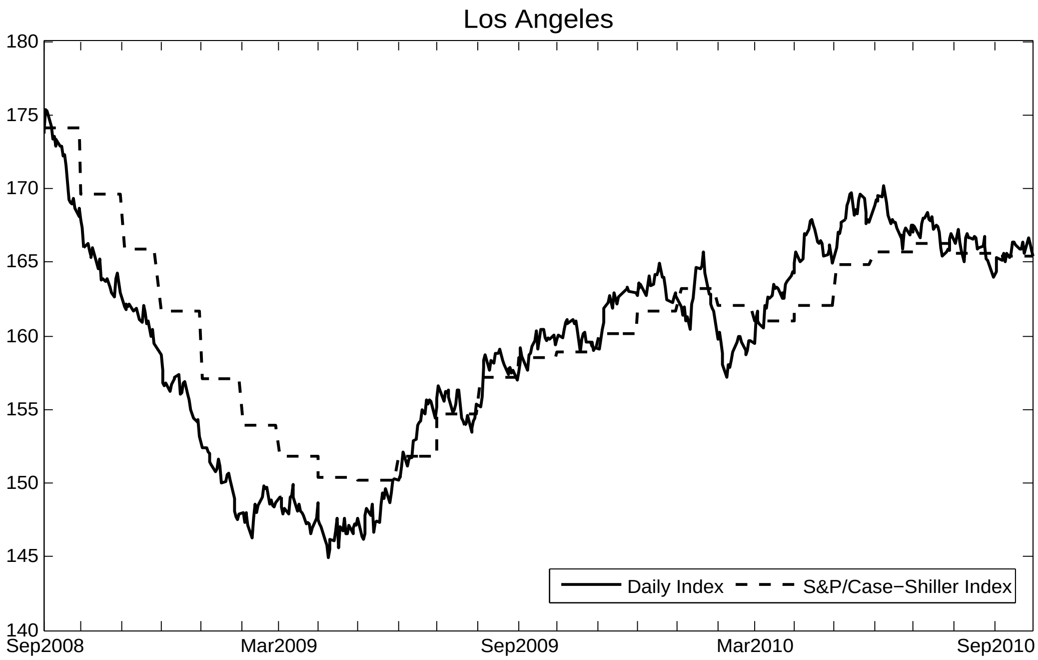
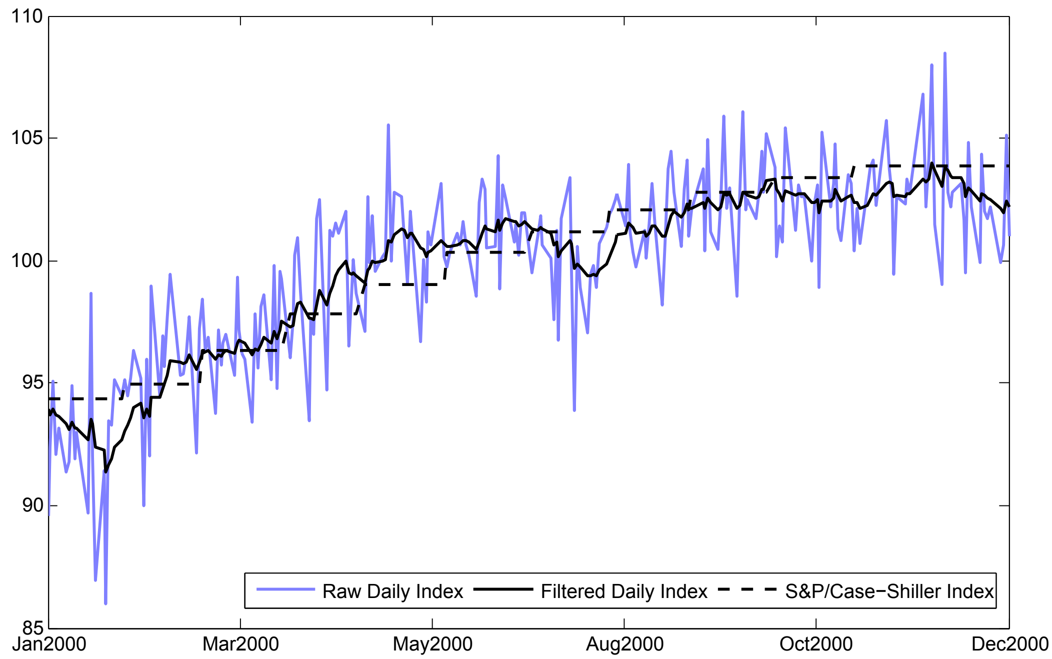
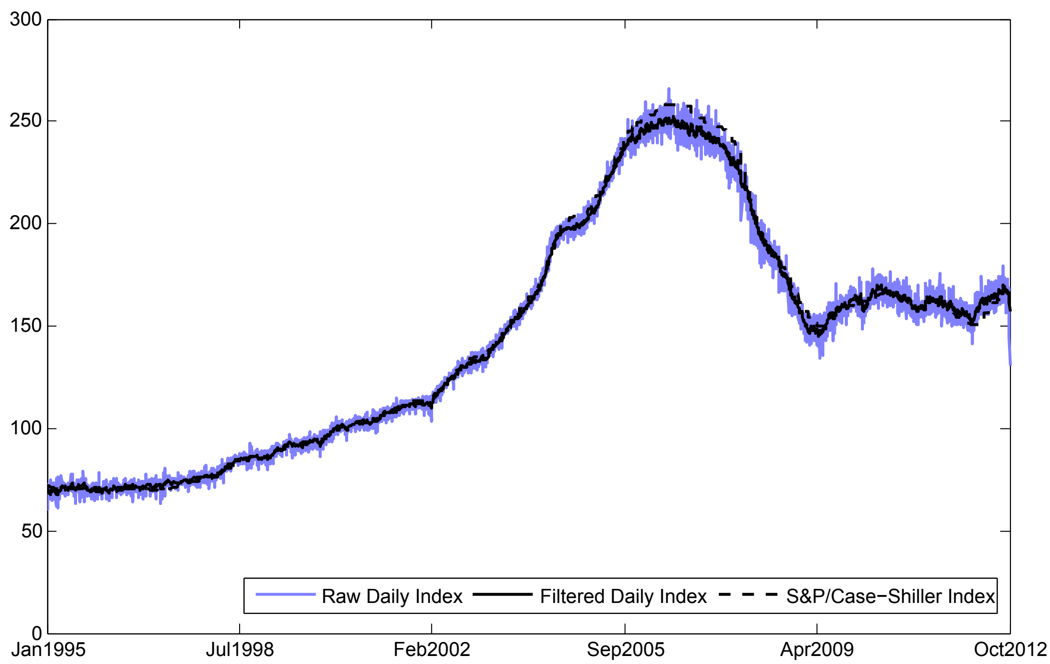


FIGURE 3.1: Daily and monthly house price indexes for Los Angeles

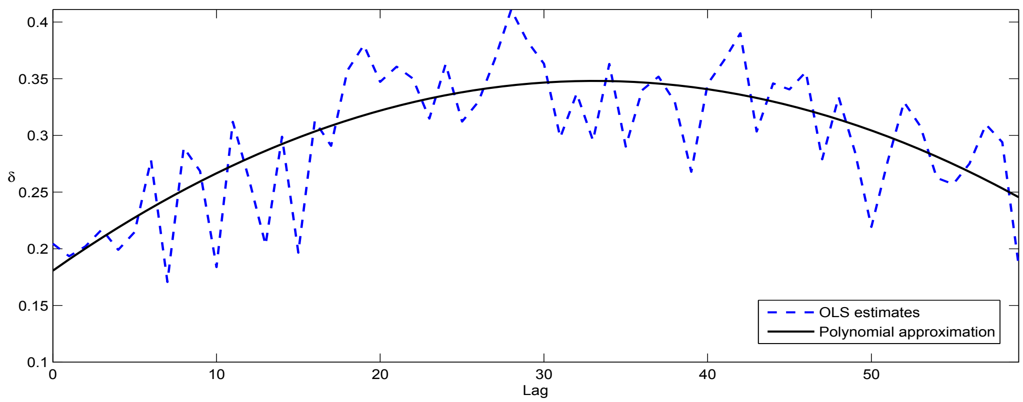


(a) January 3, 2000 to December 29, 2000

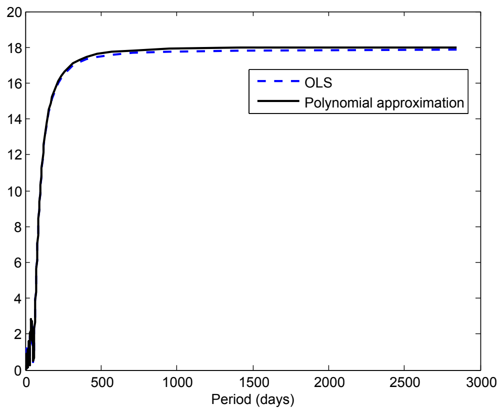


(b) January 3, 1995 to October 23, 2012

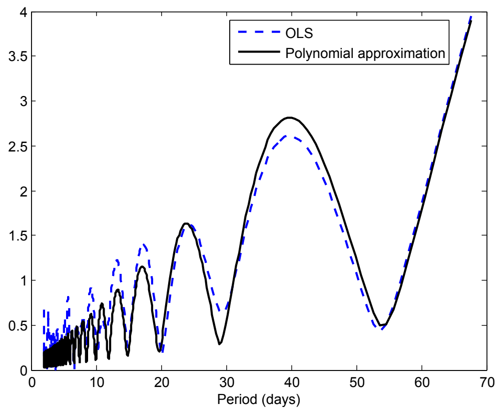
FIGURE 3.2: Raw and filtered daily house price indexes for Los Angeles



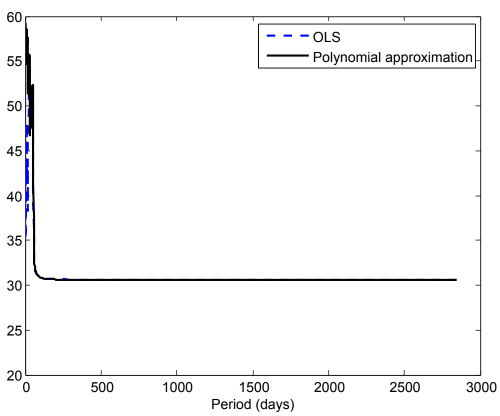
(a) Estimated $\delta(L)$ filter coefficients



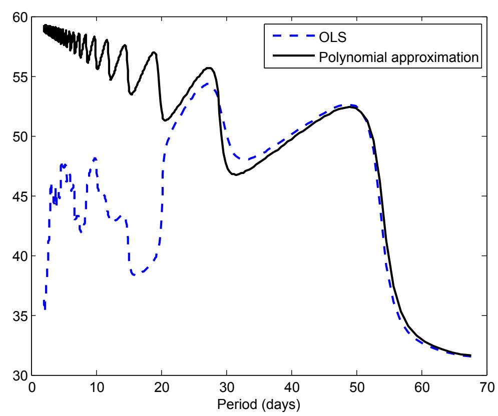
(b) Gain (all periods)



(c) Gain (shorter-run periodicities)



(d) Shift (all periods)



(e) Shift (shorter-run periodicities)

FIGURE 3.3: Characteristics of the $\delta(L)$ filter

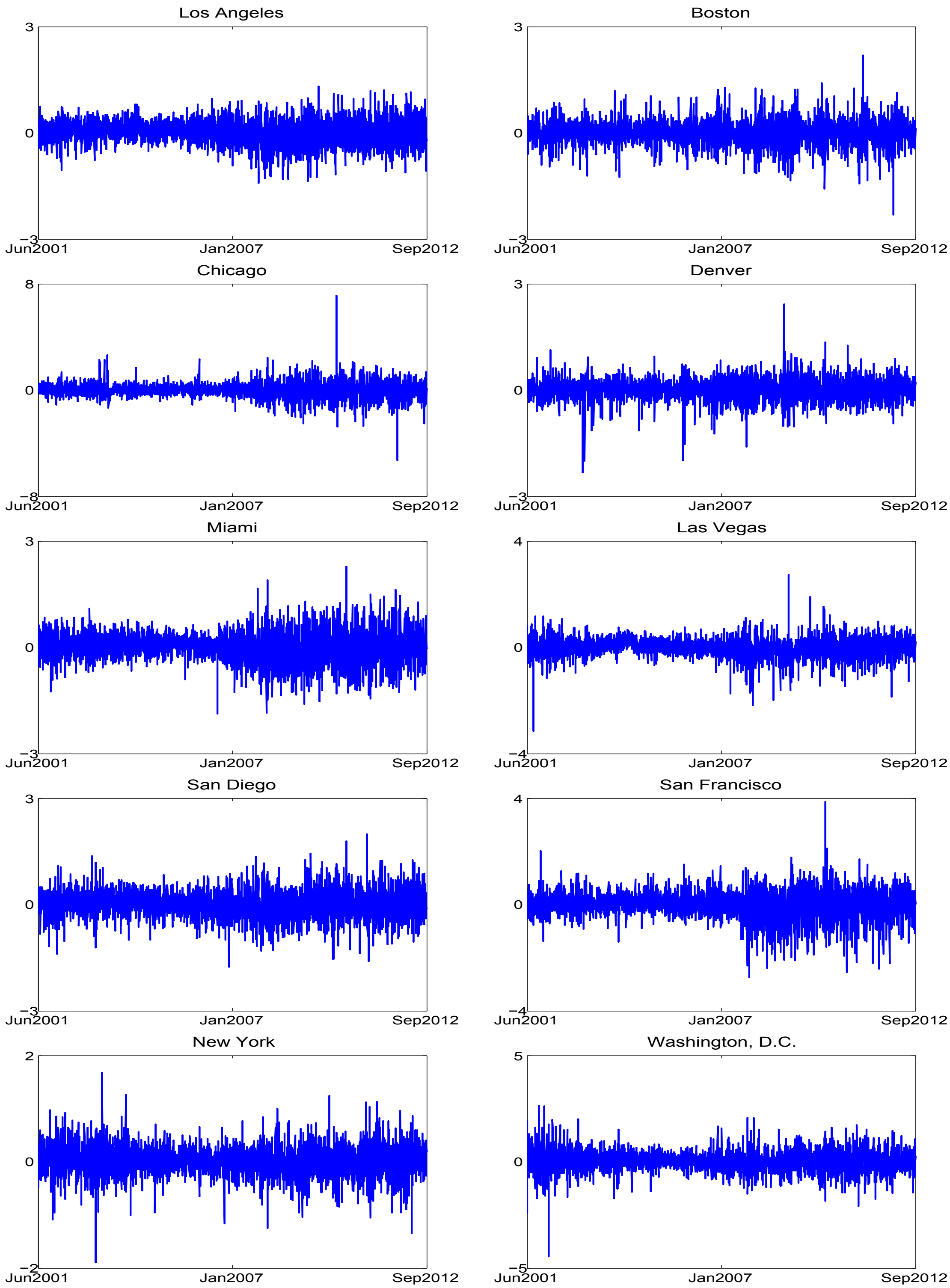


FIGURE 3.4: Daily housing returns

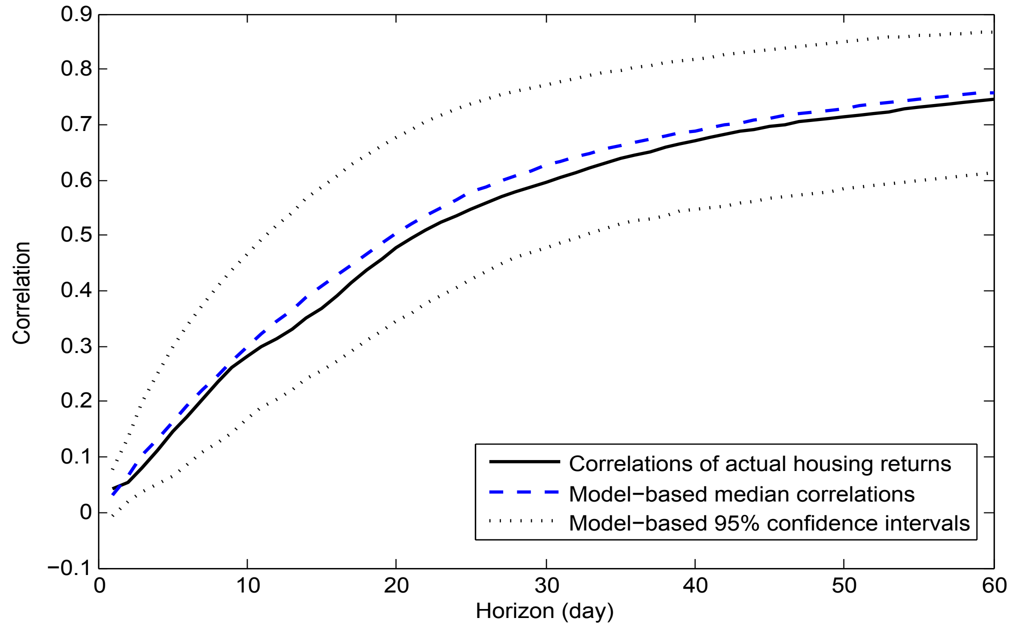


FIGURE 3.5: Unconditional return correlations for Los Angeles and New York

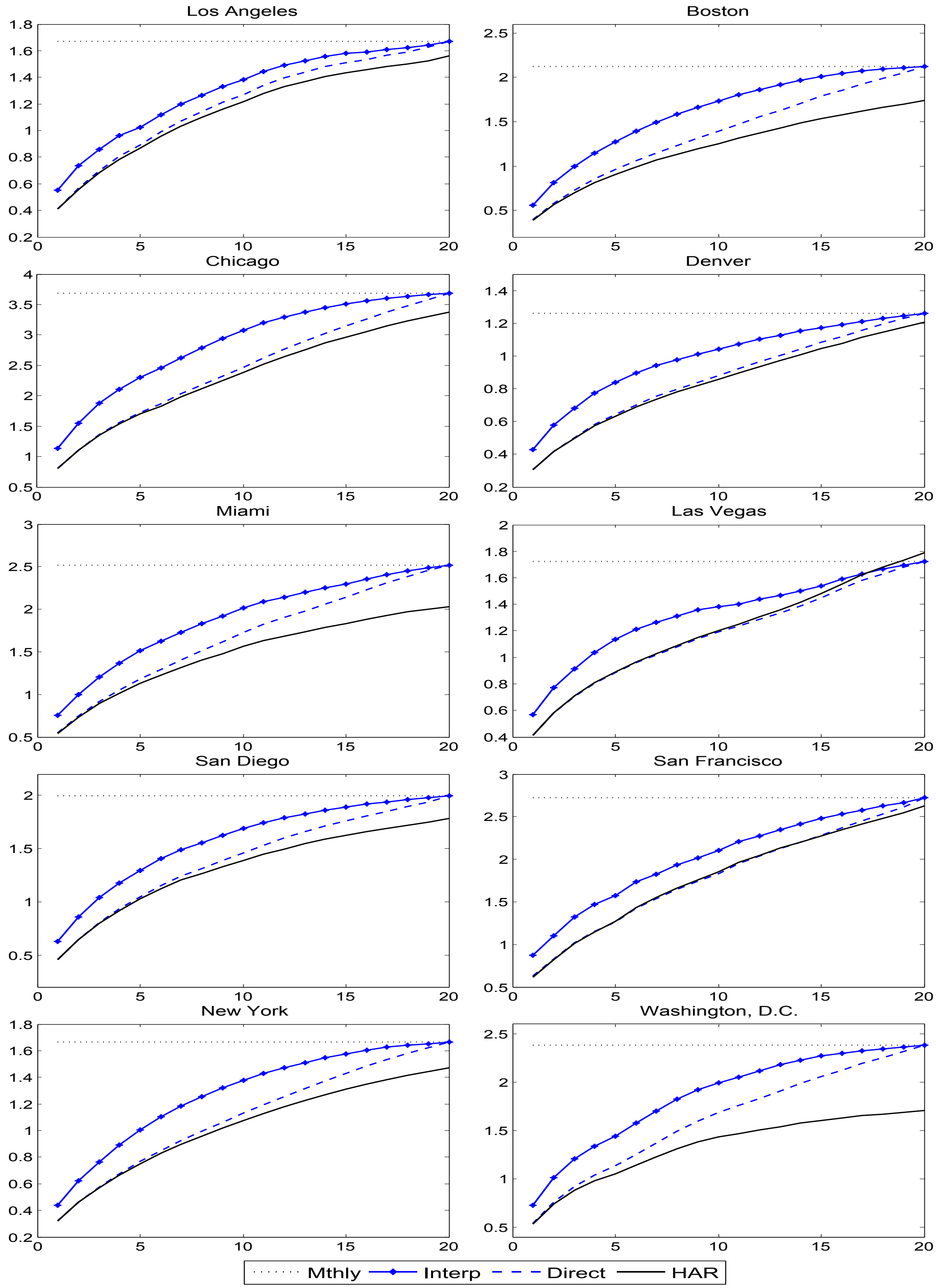


FIGURE 3.6: Forecast RMSEs as a function of forecast horizon (1 to 20 days)

4

Housing Price Volatilities: Asymmetries and Linkage to Stock Price Volatilities

4.1 Introduction

The recent economic crisis, which arguably originated with the precipitous drop in housing prices that began in 2006, directly underscores the important effects of changes in housing valuation on the capital markets and the overall economy. Because the level of housing prices is generally believed to be a reflection of consumer confidence, it is an important indicator of economic recovery. The housing market not only affects the consumer goods markets through wealth effect (Case et al., 2005, 2011), it also impacts the financial sector through mortgages and mortgage-backed securities (Miller and Peng, 2006), as well as investors' portfolio management activities. Indeed, many studies (e.g., Cocco, 2005; Yao and Zhang, 2005; Flavin and Yamashita, 2002, 2011) discuss households' portfolio management in the presence of housing and also document the benefits of including real estate or real estate securities in mixed-asset portfolios (e.g., Webb et al., 1988; Hoesli et al., 2004), thereby re-emphasizing the importance of understanding the dynamics of housing volatility and its relation to financial market volatility.

Financial market volatility has been more accurately measured and studied with the availability of high frequency intraday data in recent years. However, the most commonly used housing price measures in research are monthly S&P/Case-Shiller indexes and quarterly Office of Federal Housing Enterprise Oversight (OFHEO) indexes. The low frequency reporting makes real-time monitoring of housing price movements impossible; in addition, the aggregation reduces the apparent volatility, thus resulting in chronic underestimation of housing volatility. Set against this background, Bollerslev et al. (2013) construct a set of new daily house price series based on a standard repeat-sales method and a comprehensive housing transaction database, which provide more timely information about housing market movements than the existing monthly S&P/Case-Shiller indexes do. The higher-frequency house price indexes not only afford a more accurate measure of housing volatilities, but also shed new light on the housing volatility dynamics at the daily and weekly frequencies.

The asymmetric response of aggregate stock market volatility to negative and positive returns has been extensively documented in the literature. Recent studies have found that volatility asymmetry is more closely related to the time-varying risk premium than to the degree of financial leverage. Tauchen (2011) and Bollerslev et al. (2012) provide formal risk-based explanations though general equilibrium models that endogenously generate volatility asymmetry. In this paper, similar to that of stock volatility, negative and statistically significant relationships are found between lagged aggregate housing return and volatilities both from a set of predictive regressions based on realized variances and from GJR-GARCH type models. Although housing is a highly leveraged asset class, a direct test of housing volatilities on changes in loan-to-value ratio indicates that the observed asymmetry in housing volatility does not stem from changes in the degree of underlying housing financial leverage, but instead, may result from the risk premium carried by housing volatility.

A risk-based illustration of the economic mechanism that underlies the observed volatility asymmetry is provided through a stylized consumption-based asset pricing model with housing, which also endogenously generates a positively relationship between housing and stock volatilities. This relationship is empirically assessed in this paper through a set of predictive regressions based on the augmented Heterogeneous Autoregressive model of Realized Volatility (HAR-RV) as in Corsi (2009). The housing market and stock market volatilities are shown to move in the same direction, meaning that an increase (decrease) in the realized variance of one market will be followed by an increase (decrease) in the realized variance of the other.

In addition to housing and stock volatilities, cross-sectional return dispersions in housing and stock markets are also examined in this paper. Cross-sectional dispersion of stock returns is related to idiosyncratic volatility, which has received considerable attention in the finance literature (e.g., Campbell et al., 2001; Garcia et al., 2011) and has been shown to fluctuate with economic conditions. Our cross-sectional dispersion of housing returns is measured across houses rather than across geographic areas, using detailed transactions of every houses; therefore, it more accurately reflects how differently economic shocks affect the values of individual houses. In this paper, the cross-sectional dispersions are shown to be useful in the prediction of both within-market and cross-market realized volatilities. In particular, adding the cross-sectional dispersion of housing returns to the HAR-RV model leads to significantly better out-of-sample forecasts for daily, weekly, and monthly realized volatilities of the housing market.

The rest of the paper is organized as follows. The next section discusses the related literature. Section 4.3 gives a brief description of the datasets used in this paper. Section 4.4 examines the asymmetric effects in housing volatilities at both aggregate and individual levels, and provides a theoretical asset pricing model to illustrate the economic mechanism that underlies the empirical findings. Section

4.5 explores the volatility linkage between housing and stock markets. Section 4.6 provides further analysis of the relationship between volatility and cross-sectional dispersion of the two markets. Section 4.7 concludes and discusses related future directions.

4.2 Related Literature

One of the striking characteristics of the stock market is that aggregate market volatility responds asymmetrically to negative and positive returns. There are two leading explanations concerning the fundamental causes behind this observed volatility asymmetry. Early studies by Black (1976) and Christie (1982) attribute it to changes in balance sheet leverage, whereas others, such as Campbell and Hentschel (1992), explain it using a time-varying risk premium or volatility feedback effect. Furthermore, the leverage effect is often found to be larger for aggregate market index return than for individual stocks (see, e.g., Tauchen et al., 1996; Andersen et al., 2001.), which counters the leverage-based explanation. Most recent studies have agreed that volatility asymmetry has little to do with the underlying degree of leverage, and the self-contained general equilibrium models developed in Tauchen (2011) and Bollerslev et al. (2012) have provided a formal risk-based economic mechanism that underlies the observed lead-lag relationship of return and volatility. For the housing market, most studies, including Lamont and Stein (1999), have focused on the influence of collateralized borrowing upon the housing price dynamics, an influence that concerns the true "financial leverage" effect. Less attention has been paid to the asymmetric responses of housing volatility to negative and positive housing returns. In this paper, the housing volatility asymmetry for both aggregate market and individual areas will be examined.

The volatility linkage between housing and stock market is closely related to the extensive literature on volatility transmission. Over the past decade, most research has focused on the spillover effect among different geographic markets for the same asset class, for example equity, bonds, and foreign exchange (see, e.g., King and Wadhwani, 1990; Hamao et al., 1990; Baillie et al., 1993; Susmel and Engle, 1994.). The recent turbulence in the U.S. housing market has brought attention to the volatility spillover among local housing markets and among real estate derivatives markets (e.g., Michayluk et al., 2006; Miao et al., 2011; Hoesli and Reka, 2011; Zhu et al., 2012). Much less attention has been paid to the volatility linkage between financial markets and real sectors, an exception being the stream of studies on the relationship between the equity and oil markets (Malik and Ewing, 2009; El Hedi Arouri et al., 2011). The lack of studies on the volatility linkage between the equity and housing markets is partly due to the lack of housing price measures that are at a comparable measurement frequency to equity data. The housing price measures are usually computed and published monthly or quarterly, while high frequency intraday stock market data has been commonly used in the stochastic volatility literature in recent years. This paper makes use of the volatility measures based on the higher-frequency daily housing prices provided by Bollerslev et al. (2013) to investigate the volatility relationship between housing and stock markets.

The cross-sectional dispersion in the stock market relates to the idiosyncratic volatility that has been studied in the recent finance literature. A study by Garcia et al. (2011) formally shows that the cross-sectional dispersion is a model-free measure of average idiosyncratic variance. The idiosyncratic volatility of the stock market is found to vary with economic conditions, and it also moves with stock market volatility together countercyclically (see, e.g. Campbell et al., 2001; Stivers, 2003; Connolly and Stivers, 2006). In terms of housing cross-sectional dispersion, Plazzi et al. (2008) documented that the cross-sectional dispersion of commercial real estate returns fluctuates with macroeconomic variables that are closely related to business cycles, such as term and credit spread, inflation, and short-term interest rates. By contrast, most measures of housing return dispersion, such as in Plazzi et al. (2008) and Van Nieuwerburgh and Weill (2010), are computed across geographic areas. Because we had access to detailed transaction records of individual houses, our measure is based on the dispersion across the returns of specific houses, which means that it is a more accurate measure of the housing risk faced by a typical homeowner. Similar to the cross-sectional dispersion of the stock market, the cross-sectional dispersion of housing returns may contain information about prevailing economic conditions, so it could have predictive power for the time-series volatility of housing and stock market, and our empirical results confirm this conjecture.

4.3 Data

4.3.1 Housing transaction data

Our housing transaction data was obtained from DataQuick, a property information company. This database contains detailed transactions of more than one hundred million properties in the United States. For most of the areas, the historical transaction records extend from the late 1990s to 2012, with transactions recorded as far back as 1987 for some large metropolitan areas, such as Boston and New York. Properties are uniquely identified by property IDs. For each property transfer, the transaction date, transaction value, and loan amount are recorded. A detailed description of the data is given in Bollerslev et al. (2013).

4.3.2 Daily house price indexes

The daily house price indexes are based on the standard repeat sales model of Shiller (1991) and are constructed using the housing transaction data from DataQuick. The daily price series cover 10 major U.S. Metropolitan Statistical Areas (MSAs): Los Angeles, Boston, Chicago, Denver, Miami, Las Vegas, San Diego, San Francisco, New York and Washington, D.C., which are the same 10 MSAs in the S&P/Case-

Shiller monthly Composite 10 Index. Bollerslev et al. (2013) provide the detailed index construction procedure, as well as a description of 10 cities' daily indexes and returns. As shown in Bollerslev et al. (2013), the new higher-frequency housing price series exhibit many characteristics similar to those of aggregate financial asset price indexes, such as mild mean predictability and strong volatility clustering. In addition, the daily indexes contain more information than the traditional monthly or quarterly indexes. In this paper, the daily Composite 10 Housing Index ($P_{c,t} = \sum_{i=1}^{10} w_i \; P_{i,t}$) is used as a proxy for the aggregate (national) housing price level, which is computed as the weighted average[1] of indexes of 10 MSAs.

4.3.3 Stock data

The daily S&P500 index and daily returns for individual stocks are from the Center for Research in Security Prices (CRSP). The realized variance measures of stock returns are based on high frequency 5-min prices of the S&P500 futures contract. The sample period is chosen to be the common period of all data sets used in this paper. The shortest daily housing price index begins in June 2001, and the high frequency stock data extend from January 1993 to December 2011, so the sample period ranges from June 2001 to December 2011, a period that covers the recent boom-bust cycle of the U.S. housing market as well as the global financial crisis that occurred during this time.

The daily Composite 10 House Price Index, the S&P500 index and their returns ($r_{c,t} = \log P_{c,t} - \log P_{c,t-1}$, $r_{sp,t} = \log P_{sp,t} - \log P_{sp,t-1}$) are shown in Figure 4.1. Both housing and stock markets experienced rapid growth in the early 2000s and sudden, dramatic price declines from 2007 to 2009. Although the stock market has rebounded

[1] The weights (w_i) are identical to the ones used in the construction of the monthly Composite 10 S&P/Case-Shiller index. The specific values for each of the 10 MSAs are 0.212, 0.074, 0.089, 0.037, 0.050, 0.015, 0.055, 0.118, 0.272, and 0.078, respectively, representing the total aggregate value of the housing stock in the 10 MSAs in the year 2000.

after the recent crisis, housing prices have oscillated around 2003 level. Housing and stock returns both show evidence of volatility clustering, and the magnitude of daily stock returns is almost 10 times larger than that of daily housing returns.

4.4 Volatility Asymmetries in Housing Market

Bollerslev et al. (2013) show that similar to daily financial asset returns, the conditional volatility of daily housing returns is time-varying, highly persistent, and mean-reverting. In this paper, a detailed investigation of the volatility dynamics of housing returns is provided with a particular focus on the asymmetric return-volatility relationship, a well-known characteristic of equity market. There are two leading explanations for volatility asymmetries in the equity market. One, which attributes the asymmetric return-volatility relationship to changes in financial leverage, originated the name "leverage" effect. Following the literature, the negative correlation between lagged return and volatility is referred as "leverage effect", although it may have nothing to do with the underlying leverage. The other explanation rests on a time-varying risk premium, which is often referred as the "volatility feedback effect." Both effects could explain why housing, as a highly leveraged investment asset class, can bear volatility asymmetry. Residential housing is highly leveraged, with average loan-to-value (LTV) ratios[2] that range from 0.577 (New York) to 0.858 (Boston) for 10 MSAs; therefore, negative housing returns will increase the degree of housing leverage and thereby will also increase subsequent housing volatilities. Housing is also the largest asset held by most U.S. households, which suggests that an increase in housing volatility would raise the required rate of housing return through a decline in current price, to allow for higher future returns. Although these two explanations are characterized by opposite causality relationships, they are often

[2] The loan-to-value ratio at purchase is the ratio of the total amount of loans to the transaction value of the house at the time of the transaction. The average LTV is the sample mean of all LTVs for housing transactions in each area.

empirically difficult to distinguish in lower-frequency data.[3]

4.4.1 Realized volatility measures

The empirical examinations of volatility dynamics in this paper are based on realized variance measures. The stock realized variances are constructed using high frequency 5-min prices of the S&P500 futures contract[4]. The 5-min price observations leave a total of 77 5-min returns and one overnight return per trading day[5].

The daily realized variance (RV_t) is the sum of 78 within-day 5-min squared returns. The averages of daily realized variances of the past 5 days and 22 days, denoted as $\overline{RV}_{t-4,t}$ and $\overline{RV}_{t-21,t}$, are used as the weekly and monthly realized variances.

$$RV_t = \sum_{j=0}^{M-1} r^2_{t-j\cdot\triangle} \tag{4.1}$$

$$\overline{RV}_{t-4,t} = \frac{1}{5}\sum_{j=0}^{4} RV_{t-j} \tag{4.2}$$

$$\overline{RV}_{t-21,t} = \frac{1}{22}\sum_{j=0}^{21} RV_{t-j} \tag{4.3}$$

where $M = 78$ and $\triangle = 1/78$. It is well-demonstrated in the literature (see, e.g., Andersen et al., 2001; Barndorff-Nielsen, 2002) that the "model free" realized variance measure based on high frequency intraday data provides more accurate ex-post observations of the true return variation than the traditional sample variance based on coarser frequency returns does. Analogous to the realized variances for the stock

[3] Bollerslev et al. (2006) use high-frequency 5-minute S&P 500 future returns to differentiate between the two competing effects and find a highly significant prolonged leverage effect as well as a strong instantaneous volatility feedback effect.

[4] The choice of 5-min sampling frequency allows a reasonable balance between market microstructure effects when sampling too frequently and loss of price movement information when sampling coarsely; see, for example, the discussion in Andersen et al. (2011a).

[5] Some papers in the stochastic volatility literature exclude the overnight return from realized variances, because it exhibits different dynamics than those of 5-min returns. Similar results to those in this paper are obtained if overnight returns are excluded.

market, the daily housing realized variance HRV_t is simply the squared daily return of the Composite 10 Index. It is noteworthy that although the squared daily return is a much noisier variance measure than the realized variance based on high-frequency intraday data, the realized variances based on daily housing returns should afford more accurate ex post observations of housing return variances than those from lower-frequency monthly or quarterly housing returns.

$$HRV_t = r_{c,t}^2 \tag{4.4}$$

Figure 4.2a and 4.2b present the annualized monthly housing and stock realized volatilities (standard deviations) for each day from July 2001 to December 2011. The average monthly realized volatility of stock is about 18.726 percent, while that of housing is only 2.625 percent, which indicates that the stock market is on average 7 times more volatile than the housing market. As found in Andersen et al. (2001), the logarithmic transformation renders the stock realized variance approximately normally distributed. The descriptive statistics shown in Table 4.2 indicate that the unconditional distribution of both realized variances are skewed and leptokurtic, whereas the logarithmic transformation significantly reduces the excess kurtosis and skewness and makes the variance measures approximately Gaussian. Therefore, logarithmic variance measures are used in the predictive regressions.

4.4.2 Volatility asymmetries in predictive regressions

Guided by the empirical literature on modeling equity market realized variances, here I rely on the Heterogeneous Autoregressive model of Realized Volatility (HAR-RV), which is a simple and parsimonious way of capturing the characteristics of realized volatility, such as long memory and fat tails. The HAR-RV model was originally proposed by Corsi (2009), and has been successfully employed in closely related contexts by including additional terms, either in the original model or in the logarithmic form[6].

The HAR-RV model is augmented by the term, capturing the additional effect of negative lagged returns on volatility.

$$\begin{aligned}\log(\overline{HRV}_{t+1,t+\tau}) &= c_\tau + \sum_{j=1}^{3} \beta_{j,\tau}\mathbf{1}_{\{t+\tau\in Q_j\}} + \beta_\tau^{(d)}\log(HRV_t) + \beta_\tau^{(w)}\log(\overline{HRV}_{t-4,t}) \\ &+ \beta_\tau^{(m)}\log(\overline{HRV}_{t-21,t}) + \gamma_\tau \log(HRV_t)\mathbf{1}_{\{r_{c,t}<0\}} + \omega_{t+\tau} \qquad (4.5)\\ \log(\overline{RV}_{t+1,t+\tau}) &= c_\tau + \sum_{j=1}^{3} \beta_{j,\tau}\mathbf{1}_{\{t+\tau\in Q_j\}} + \beta_\tau^{(d)}\log(RV_t) + \beta_\tau^{(w)}\log(\overline{RV}_{t-4,t}) \\ &+ \beta_\tau^{(m)}\log(\overline{RV}_{t-21,t}) + \gamma_\tau \log(RV_t)\mathbf{1}_{\{r_{sp,t}<0\}} + \omega_{t+\tau} \qquad (4.6)\end{aligned}$$

where $\mathbf{1}_{\{t+\tau\in Q_j\}}$ indicates whether $t+\tau$ belongs to the jth quarter, which controls for potential seasonality in realized variances[7]. τ is the forecast horizon, in which $\tau = 1$, $\tau = 5$ and $\tau = 22$ correspond to (approximately) daily, weekly, and monthly horizons, respectively. $\overline{HRV}_{t+1,t+\tau} = \frac{1}{\tau}\sum_{j=1}^{\tau} HRV_{t+j}$ and $\overline{RV}_{t+1,t+\tau} = \frac{1}{\tau}\sum_{j=1}^{\tau} RV_{t+j}$ are average housing and stock realized variances from day $t+1$ to $t+\tau$. If γ_τ is different from zero under this specification, then volatility asymmetry exists. The logarithmic daily realized variances ($\log(HRV_t)$, $\log(RV_t)$)) are all negative throughout the sample, so γ_τ should be negative to capture the fact that volatility usually increases following bad news. The estimation results are shown in Table 4.7. For both markets, the asymmetry coefficients γ_τ are small in magnitude. For the stock market, γ_τ is negatively significant for all daily, weekly and monthly forecast horizons, which is in line with the findings documented in the literature (e.g., Patton and Sheppard, 2011). Meanwhile, for the housing market, γ_τ is negative, but not statis-

[6] For this stream of literature on empirical modeling of realized volatility measures, see, e.g., Andersen et al. (2007), Bollerslev et al. (2009) and Andersen et al. (2011b).

[7] If monthly dummies are used to control for seasonality, all regression results are similar to those reported here, and the realized variances for the housing market in January turn out to be significantly higher than those in other months.

tically significant at the daily and weekly horizon; it becomes significantly negative as the horizon increases to monthly. The lack of evidence for volatility asymmetry for housing at short horizons is not surprising, since the squared daily return is a much noisier variance measure than one constructed from high-frequency 5 min returns. This observation is also based on the fact that R^2s of predictive regressions for housing realized variance are much lower than those from stock realized variance regressions, especially at short forecast horizons.

One interesting finding in the literature is that although volatility asymmetry is found for aggregate stock returns, it tends to be weaker at the individual stock level (see, e.g. Tauchen et al., 1996; Andersen et al., 2001); this evidence is contradictory to the leverage-based explanation. Similarly, volatility asymmetry found in aggregate housing market does not necessarily exist in individual MSAs. The same predictive regressions can be implemented for each MSA in order to investigate the presence of volatility asymmetries in individual housing markets.

$$\begin{aligned} \log(\overline{HRV}_{i,t+1,t+\tau}) &= c_{i,\tau} + \sum_{j=1}^{3} \beta_{i,j,\tau} \mathbf{1}_{\{t+\tau \in Q_j\}} + \beta_{i,\tau}^{(d)} \log(HRV_{i,t}) \\ &+ \beta_{i,\tau}^{(w)} \log(\overline{HRV}_{i,t-4,t}) + \beta_{i,\tau}^{(m)} \log(\overline{HRV}_{i,t-21,t}) \\ &+ \gamma_{i,\tau} \log(HRV_{i,t}) \mathbf{1}_{\{r_{i,t}<0\}} + \omega_{i,t+\tau} \end{aligned} \tag{4.7}$$

The estimation results are reported in Table 4.3. γ_τ is negative and statistically significant at the monthly forecast horizon for seven (Los Angeles, Boston, Denver, Miami, Las Vegas, San Diego and New York) out of ten MSAs. The only exception is New York, where evidence of volatility asymmetry is shown at the weekly horizon.

4.4.3 Volatility asymmetries and degree of financial leverage

Although it is unlikely that the volatility asymmetries in housing markets found at daily frequency stem from the corresponding changes in the degree of financial leverage, it is still instructive to formally test this hypothesis before turning to other explanations. A direct measure of the degree of financial leverage for housing is the loan-to-value (LTV) ratio, which is the proportion of loans (usually secured by the property) in relation to its transaction value. Let $LTV_{i,t}$ be the average loan-to-value ratio[8] of all transactions in area i at day t. In the predictive regression, suppose the original asymmetry parameter is a function of the daily changes in the loan-to-value ratio ($\triangle LTV_{i,t}$) as follows.

$$\begin{aligned}
\log(\overline{HRV}_{i,t+1,t+\tau}) &= c_{i,\tau} + \sum_{j=1}^{3} \beta_{i,j,\tau} \mathbf{1}_{\{t+\tau \in Q_j\}} + \beta_{i,\tau}^{(d)} \log(HRV_{i,t}) \\
&+ \beta_{i,\tau}^{(w)} \log(\overline{HRV}_{i,t-4,t}) + \beta_{i,\tau}^{(m)} \log(\overline{HRV}_{i,t-21,t}) \\
&+ (\gamma_{i,\tau} + \gamma_{i,\tau}^{'} \triangle LTV_{i,t}) \log(HRV_{i,t}) \mathbf{1}_{\{r_{i,t}<0\}} + \omega_{i,t+\tau} \qquad (4.8)
\end{aligned}$$

where $\gamma_{i,\tau}^{'}$ capture the effect of changes in loan-to-value ratio on the future housing realized variance, when the current housing return is negative. If the observed volatility asymmetry is not related to the changes in LTV, $\gamma_{i,\tau}^{'}$ should be insignificant. It can be seen from Table 4.4 that $\gamma_{i,\tau}^{'}$ is generally not statistically different than zero for all horizons, and that $\gamma_{i,\tau}$ is not affected after controlling for changes in LTV for all MSAs. This result suggests that the volatility asymmetries at individual housing markets, although they are relatively weak, do not result from changes in degree of the underlying financial leverage. A similar test can be done for the realized variances of the aggregate housing market.

$$\begin{aligned}
\log(\overline{HRV}_{t+1,t+\tau}) &= c_{\tau} + \sum_{j=1}^{3} \beta_{j,\tau} \mathbf{1}_{\{t+\tau \in Q_j\}} + \beta_{\tau}^{(d)} \log(HRV_{t}) \\
&+ +\beta_{\tau}^{(w)} \log(\overline{HRV}_{t-4,t}) + \beta_{\tau}^{(m)} \log(\overline{HRV}_{t-21,t}) \\
&+ (\gamma_{\tau} + \gamma_{\tau}^{'} \triangle LTV_{t}) \log(HRV_{t}) \mathbf{1}_{\{r_{c,t}<0\}} + \omega_{t+\tau} \qquad (4.9)
\end{aligned}$$

[8] The ratio used here is a combined loan-to-value, which includes all (up to three) loans on the property besides the primary loan.

where LTV_t is a weighted average of all $LTV_{i,t}$, using the same weights as in the daily Composite 10 Index. Similar to the results for individual MSAs, γ'_τ is not statistically significant for all horizons, and γ_τ is still negative and statistically significant after controlling for changes in LTV. Overall, these results indicate that the observed asymmetric responses of housing volatility to negative and positive returns have little connection to the underlying degree of housing leverage; this indication is similar to the findings in Figlewski and Wang (2000), which show that the asymmetric effects in stock volatilities have little to do with the underlying firm leverages.

4.4.4 Robustness check

In addition to the predictive regressions, the GARCH type of models, including E-GARCH and GJR-GARCH, are well-established as formal tests of the asymmetric effect in volatility at daily frequency. As a robustness check of the results from predictive regressions, I employed the GJR-GARCH model proposed by Glosten et al. (1993) to examine the volatility asymmetries in the housing market at both aggregate and individual levels[9].

For housing returns in individual areas, Bollerslev et al. (2013) propose a HAR-X-GARCH model and show that it fits daily returns for all 10 MSAs reasonably well.[10] For the conditional mean part of the model, the own fifth lag of the returns is included here to account for any weekly calendar effects, and the own monthly and composite monthly returns provide a parsimonious way of accounting for longer-run city-specific and common national dynamic dependencies. Asymmetric responses to positive and negative shocks are allowed in the conditional variance structure as in

[9] Engle and Ng (1993) compare various GARCH type models that could test leverage effect and find that the GJR model is the best parametric one.

[10] Bollerslev et al. (2013) provide detailed specification tests of the HAR-X-GARCH model.

the GJR-GARCH specification.

$$
\begin{aligned}
r_{i,t} &= \mu_i + \rho_{i1} r_{i,t-1} + \rho_{i5} r_{i,t-5} + \rho_{im} r^m_{i,t-1} + \rho_{ic} r^m_{c,t-1} + \varepsilon_{i,t} \qquad (4.10)\\
\varepsilon_{i,t}|\Omega_{t-1} &\sim N(0, h_{i,t})\\
Var_{t-1}(r_{i,t}) &\equiv h_{i,t} = \omega_i + \alpha_i \varepsilon^2_{i,t-1} + \beta_i h_{i,t-1} + \gamma_i \varepsilon^2_{i,t-1} \mathbf{1}_{\{\varepsilon_{i,t-1}<0\}} \qquad (4.11)
\end{aligned}
$$

All parameters are estimated simultaneously by the standard (quasi-)maximum likelihood method. For each area, the estimation results are shown in Table 4.5 with robust standard errors as in Bollerslev and Wooldridge (1992). The small and insignificant values of γ for all 10 MSAs indicate that volatility asymmetry is not found at the individual MSA level. The lack of evidence for volatility asymmetry in individual housing markets is in line with the fact that it is also not found at daily horizon from the predictive regressions.

For the aggregate housing market, I consider a similar model for returns of the Composite 10 Index. Instead of using only the first and fifth lags in the conditional mean equation for returns of individual areas, I include all five lagged returns for modeling the conditional mean, because the composite return is more persistent than the individual series[11]. I also rely on the GJR-GARCH specification to examine the asymmetric effect in aggregate market volatility.

$$
\begin{aligned}
r_{c,t} &= \mu_c + \sum_{j=1}^{5} \rho_j \ r_{c,t-j} + \rho_m \ r^m_{c,t-1} + \varepsilon_{c,t} \qquad (4.12)\\
\varepsilon_{c,t}|\Omega_{t-1} &\sim N(0, h_{c,t})\\
Var_{t-1}(r_{c,t}) &\equiv h_{c,t} = \omega + \alpha \ \varepsilon^2_{c,t-1} + \beta \ h_{c,t-1} + \gamma \ \varepsilon^2_{c,t-1} \mathbf{1}_{\{\varepsilon_{c,t-1}<0\}} \qquad (4.13)
\end{aligned}
$$

The (quasi-)maximum likelihood estimates are shown in Table 4.6. The estimated γ is positive and statistically significant, although the magnitude of 0.036 is relatively small. Positive γ under the GJR-GARCH specification suggests that volatility

[11] The volatility asymmetry result is robust to the reasonable choices of the number of lagged returns, and is also robust to the GARCH-in-mean specification.

will increase more following a negative return shock, which is evidence of volatility asymmetry for daily aggregate housing returns. For comparison, I next consider a simple GJR-GARCH model for S&P 500 index returns($r_{sp,t}$). The well-documented volatility asymmetry of the stock market is found for this sample period; similar to the housing asymmetry parameter, γ has a relatively small magnitude of 0.137.

$$r_{sp,t} = \mu_{sp} + \rho_1 \; r_{sp,t-1} + \varepsilon_{sp,t} \tag{4.14}$$

$$\varepsilon_{sp,t}|\Omega_{t-1} \sim N(0, h_{sp,t})$$

$$Var_{t-1}(r_{sp,t}) \equiv h_{sp,t} = \omega + \alpha \; \varepsilon^2_{sp,t-1} + \beta \; h_{sp,t-1} + \gamma \; \varepsilon^2_{sp,t-1}\mathbf{1}_{\{\varepsilon_{sp,t-1}<0\}} \tag{4.15}$$

Overall, strong evidence of volatility asymmetry exists for both the aggregate housing and stock markets. This result is in line with the results from predictive regressions, although the noisy housing realized variances render the asymmetry parameters statistically insignificant at short horizons.

Although the GJR-GARCH model and the predictive regressions reach similar conclusions on the volatility asymmetries for the aggregate and individual housing markets, the regression approach has a clear advantage over the GJR-GARCH approach. The GJR-GARCH model, which is mostly used to analyze the return-volatility relationship at daily frequency, can hardly give accurate estimates at monthly frequency, if the sample size is not big enough. The predictive regression provides a simple and accurate way to gauge the return-volatility relationship at daily, weekly, and monthly horizons.

4.4.5 Volatility asymmetries in equilibrium

Tauchen (2011) develops a general equilibrium model that endogenously generates a dynamic leverage effect in the stock market, the sign of which depends directly on the coefficient of risk aversion and the inter-temporal elasticity of substitution of a representative agent. Bollerslev et al. (2012) further provide an equilibrium model that can explain not only volatility asymmetry, but also other empirical facts of the dynamic dependencies of stock market volatility, such as long memory volatility and short memory volatility risk premium. Moreover, their model is cast in continuous time to avoid the assumption of the agent's decision interval.

As a stylized illustration of the mechanism of leverage effects in the housing market, this paper follows the discrete time framework in Tauchen (2011) and introduce housing as an asset that pays a stream of housing services. The representative agent derives utility from a consumption bundle composed of housing service and nonhousing consumption as in Piazzesi et al. (2007). The detailed model setup and solutions are given in Appendix B. This model can endogenously generate negative correlations between lagged return and volatility for both housing and stock assets, which is in line with the empirical volatility asymmetries presented in this section. Although the model only sheds light on the economic mechanism of volatility asymmetries in aggregate housing and stock markets, it does not directly explain the evidence for volatility asymmetry in individual housing markets. Similar to the aggregate housing market volatility, housing volatility at individual MSA levels should also carry risk premiums, but the large idiosyncratic risks in individual MSAs sometimes render the volatility asymmetry parameters statistically insignificant.

Aside from the asymmetric effects, housing and stock volatilities are also endogenously linked in the model, which generates the positive correlations between them. In short, an increase (decrease) in the volatility of one market should be followed by an increase (decrease) in the volatility of the other market. The next section presents an empirical investigation into this volatility linkage between the two markets.

4.5 Linkage between Housing and Stock Volatilities

A natural way to study this conditional correlation is to examine volatility transmission through multivariate GARCH type models, which typically require accurate time matching across multivariate time series. However, unlike in equity trading, no centralized market for housing transactions monitors exact deal closure dates. Usually a time gap exists between deal closure date and transaction publicly recorded date. The length of this gap varies, but it could be as long as several weeks, during which house buyers and/or sellers usually go through loan or document preparation processes. The timing mismatch between housing and stock markets suggests that the traditional M-GARCH type models might not be good tools with which to uncover the dynamic conditional correlations of the two volatilities. Similar to the method used in Section 4.4.2, predictive regressions of realized variances of one market on those of the other can be implemented separately over different horizons; this is a simple way to solve the problem of the date gap. Before turning to the results of such predictive regressions, let us first look at the unconditional correlations of two realized variance measures over daily, weekly, and monthly horizons (Table 4.1). The correlation increases from 0.082 to 0.307 as the horizon widens from daily to monthly, which suggests that the two variance measures are more closely correlated at longer horizons. Therefore, monthly realized variances are used as predictors; moreover, monthly aggregation could also alleviate the problem of date mismatch between the two volatility series.

First, the HAR-RV model of realized variance of one market is augmented by the monthly realized variance of the other[12]. Although the unknown deal closure dates of housing transactions are always before the record dates, the transactions are only known to the public on the record dates. For this reason, the housing volatility measures are constructed from the daily returns that are based on the record dates

[12] The HAR-RV model referred to hereafter is the original HAR-RV model in Corsi (2009) in logarithmic form with an asymmetry term and quarterly dummies.

of the housing transactions.

$$
\begin{aligned}
\log(\overline{HRV}_{t+1,t+\tau}) &= c_\tau + \sum_{j=1}^{3} \beta_{j,\tau} \mathbf{1}_{\{t+\tau \in Q_j\}} + \beta_\tau^{(d)} \log(HRV_t) \\
&+ \beta_\tau^{(w)} \log(\overline{HRV}_{t-4,t}) + \beta_\tau^{(m)} \log(\overline{HRV}_{t-21,t}) \\
&+ \gamma_\tau \log(HRV_t) \mathbf{1}_{\{r_{c,t}<0\}} + \beta_\tau^{(S)} \log(\overline{RV}_{t-21,t}) + \omega_{t+\tau} \quad (4.16)
\end{aligned}
$$

$$
\begin{aligned}
\log(\overline{RV}_{t+1,t+\tau}) &= c_\tau + \sum_{j=1}^{3} \beta_{j,\tau} \mathbf{1}_{\{t+\tau \in Q_j\}} + \beta_\tau^{(d)} \log(RV_t) \\
&+ \beta_\tau^{(w)} \log(\overline{RV}_{t-4,t}) + \beta_\tau^{(m)} \log(\overline{RV}_{t-21,t}) \\
&+ \gamma_\tau \log(RV_t) \mathbf{1}_{\{r_{sp,t}<0\}} + \beta_\tau^{(H)} \log(\overline{HRV}_{t-21,t}) + \omega_{t+\tau} \quad (4.17)
\end{aligned}
$$

Table 4.7 presents the values of the estimated parameters and the corresponding heteroskedasticity and autocorrelation robust standard errors, as in Newey and West (1987)[13] with lags of $\lfloor T^{1/3} \rfloor + 2\tau$. $\beta_\tau^{(S)}$ is significantly positive for all horizons from daily to monthly, which indicates that an increase in stock market volatility will be followed by an increase of volatility in the housing market. The inclusion of $\log(\overline{RV}_{t-21,t})$ only affects the parameter magnitude of $\log(\overline{HRV}_{t-21,t})$, while leaving the estimated coefficients for $\log(HRV_t)$ and $\log(\overline{HRV}_{t-4,t})$ almost intact. The R^2s are increased by 0.2% to 4.6% from daily to monthly horizon. By including monthly housing realized variance $\log(\overline{HRV}_{t-21,t})$ in the HAR-RV model of stock realized variance as in (4.17), $\beta_h^{(H)}$ is significantly positive for the monthly forecast horizon ($\tau = 22$). This result suggests that higher housing market volatility implies higher monthly stock market volatility in the future. Also, similar to the estimation result of (4.16), adding $\log(\overline{HRV}_{t-21,t})$ to the predictive HAR-RV model of stock realized

[13] Similar conclusions from predictive regressions could be drawn in this paper if the standard errors from the data driven automatic lag selection in covariance matrix estimation are used, as in Newey and West (1994).

variance only reduces the parameter magnitude of $\log(\overline{RV}_{t-21,t})$, an effect that is in line with the strong longer-run unconditional correlation of the two realized variances. However, R^2s are not dramaticallly increased by including $\log(\overline{HRV}_{t-21,t})$; this is not surprising, considering that the standard HAR-RV model has already explained a large proportion (more than 70%) of stock realized variances.

Notice that the τ-day-ahead conditional variance forecasts can be readily obtained from the GJR-GARCH type models of daily housing and stock returns shown in Section 4.4.4. The GJR-GARCH forecasted variances could also be used to examine the volatility linkage between the two markets.

Let $\overline{HV}^{GARCH}_{t+1,t+\tau}$ and $\overline{SV}^{GARCH}_{t+1,t+\tau}$ be the average housing and stock variances from day $t+1$ to $t+\tau$, estimated from the GJR-GARCH models in Section 4.4.4,

$$\overline{HV}^{GARCH}_{t+1,t+\tau} = \frac{1}{\tau}\sum_{j=1}^{\tau}\hat{\sigma}^2_{c,t+j|t} \tag{4.18}$$

$$\overline{SV}^{GARCH}_{t+1,t+\tau} = \frac{1}{\tau}\sum_{j=1}^{\tau}\hat{\sigma}^2_{s,t+j|t} \tag{4.19}$$

where the variances of housing and stock returns at date $t+j$ conditional on the information of date t can be estimated by iterating (4.13) and (4.15) j-step forward.

$$\hat{\sigma}^2_{c,t+j|t} = \widehat{Var}_t\left[\varepsilon_{c,t+j}\right] \tag{4.20}$$

$$\hat{\sigma}^2_{s,t+j|t} = \widehat{Var}_t\left[\varepsilon_{sp,t+j}\right] \tag{4.21}$$

Next, the GJR-GARCH estimated variances are used to predict future housing and stock realized variances.

$$\begin{aligned}\log(\overline{HRV}_{t+1,t+\tau}) &= c_\tau + \sum_{j=1}^{3}\beta_{j,\tau}\mathbf{1}_{\{t+\tau\in Q_j\}} + \beta^{(d)}_\tau\log(HRV_t)\\ &+ \beta^{(w)}_\tau\log(\overline{HRV}_{t-4,t}) + \beta^{(m)}_\tau\log(\overline{HRV}_{t-21,t})\\ &+ \gamma_\tau\log(HRV_t)\mathbf{1}_{\{r_{c,t}<0\}} + \beta^{(S)}_\tau\log(\overline{SV}^{GARCH}_{t+1,t+\tau}) + \omega_{t+\tau}\end{aligned} \tag{4.22}$$

$$
\begin{aligned}
\log(\overline{RV}_{t+1,t+\tau}) &= c_\tau + \sum_{j=1}^{3} \beta_{j,\tau} \mathbf{1}_{\{t+\tau \in Q_j\}} + \beta_\tau^{(d)} \log(RV_t) \\
&+ \beta_\tau^{(w)} \log(\overline{RV}_{t-4,t}) + \beta_\tau^{(m)} log(\overline{RV}_{t-21,t}) \\
&+ \gamma_\tau \log(RV_t) \mathbf{1}_{\{r_{sp,t}<0\}} + \beta_\tau^{(H)} \log(\overline{HV}^{GARCH}_{t+1,t+\tau}) + \omega_{t+\tau} \quad (4.23)
\end{aligned}
$$

As shown in Table 4.7, these results are similar to results that use past realized variances as discussed previously in this section. The GJR-GARCH model implies that τ-days ahead conditional variance forecast for one market is of use in predicting realized variances of the other market for both weekly and monthly horizons. The fact that higher (lower) realized variance in one market is observed if higher (lower) conditional variance is expected in the other market again confirms the positive volatility relationship between the two markets.

4.6 Cross-sectional Dispersions

Many studies have found that the cross-sectional dispersion of stock returns, which is a measure of average idiosyncratic variance, is closely related to economic conditions or aggregate economic uncertainty (see, e.g., Stivers, 2003; Garcia et al., 2011), with higher (lower) dispersion in a business cycle downturn (upturn). Similarly, the cross-sectional dispersion of housing returns might also be driven by prevailing economic conditions, because the economic shocks are propagated differently not only across regions (Plazzi et al., 2008) but also across houses. Houses are heterogeneous assets, and homeowners also have different demographic characteristics. An increase in economic uncertainty increases the dispersion of some demographic characteristics, such as income due to heterogeneous abilities, and may also generate many corresponding housing activities, such as forced moves and foreclosures, both of which increase the cross-sectional dispersion of housing returns within a region[14]. Therefore, be-

[14] Van Nieuwerburgh and Weill (2010) use a spatial equilibrium model to study the house price dispersion across metropolitan areas, and argue that faced with an increase in the productivity dis-

cause the cross-sectional dispersions of housing and stock returns are closely linked to economic fluctuations, they should be positively correlated with housing and stock volatilities, which will be examined in this section.

4.6.1 Cross-sectional variance

For the housing market, let $R_{i,j,t}$ be the annualized logarithmic return of house i in area j that sold at time t, and previously sold at time s.

$$R_{i,j,t} = \frac{1}{(t-s)/365}(\log P_{i,j,t} - \log P_{i,j,s}) \tag{4.24}$$

If there are in total $N_{j,t}$ houses sold at date t in area j, the cross-sectional variance of returns at time t is the sample variance of $N_{j,t}$ annualized log returns.

$$HCV_{j,t} = \frac{1}{N_{j,t}-1}\sum_{i=1}^{N_{j,t}}(R_{i,j,t} - \bar{R}_{j,t})^2 \tag{4.25}$$

The cross-sectional variance for the housing market is the weighted average of the cross-sectional variance of all areas.

$$HCV_t = \sum_{j=1}^{10} w_j HCV_{j,t} \tag{4.26}$$

Figure 4.2c shows that the cross-sectional housing dispersion dramatically increased during the turbulent period of the housing market from late 2007 to early 2010. To compute the cross-sectional variances of the stock market, I obtained the data for 30 stocks in the Dow Jones Industrial Average from CRSP. The cross-sectional variance at each day, SCV_t, is the sample variance of daily returns across the 30 stocks. As shown in Figure 4.2b and 4.2d, the stock cross-sectional variance and stock volatility

persion across areas, households choose to reallocate from lower to higher productivity metropolitan areas, a choice that generates increases in the observed cross-sectional dispersions of both house prices and wages. The measure of housing return dispersion used in this paper is at household level, which more accurately reflects the risk faced by homeowners.

show similar countercyclical patterns, and both peaked around late 2008. It is also noteworthy that the average annualized housing cross-sectional dispersion is 12.754 percent, much higher than that of the stock market with only 1.566 percent. The large difference comes from the way the two cross-sectional variances are constructed. Unlike that for stock, the cross-sectional dispersion of housing has nothing to do with the daily housing returns. Instead, it reflects an accumulation of all the risks through the time interval (6 years on average) between two successive transactions involving the same house.

4.6.2 Cross-sectional variances and volatilities of two markets

The unconditional correlations shown in Table 4.1 indicate that the within-market correlations of volatility and cross-sectional dispersion are the highest (0.700 for the housing market and 0.682 for the stock market, at monthly frequency). The cross-market correlations between volatility and cross-sectional dispersion are also higher than correlations between the two volatilities. For example, at the monthly frequency, the correlation of HCV and RV is 0.433 and the correlation of SCV and HRV is 0.316; both are higher than the correlation of RV and HRV, which is 0.307. Similar to the methodology of studying the volatility relationship of the two markets, predictive regression is also employed to study the relationship between volatility and cross-sectional dispersion. I augment the HAR-RV model with monthly average housing cross-sectional variance.

$$\begin{aligned}
\log(\overline{HRV}_{t+1,t+\tau}) &= c_\tau + \sum_{j=1}^{3} \beta_{j,\tau} \mathbf{1}_{\{t+\tau \in Q_j\}} + \beta_\tau^{(d)} \log(HRV_t) \\
&+ \beta_\tau^{(w)} \log(\overline{HRV}_{t-4,t}) + \beta_\tau^{(m)} \log(\overline{HRV}_{t-21,t}) \\
&+ \gamma_\tau \log(HRV_t) \mathbf{1}_{\{r_{c,t}<0\}} + \beta_\tau^{(H)} \log(\overline{HCV}_{t-21,t}) + \omega_{t+\tau} \quad (4.27)
\end{aligned}$$

$$\begin{aligned}\log(\overline{RV}_{t+1,t+\tau}) &= c_\tau + \sum_{j=1}^{3}\beta_{j,\tau}\mathbf{1}_{\{t+\tau\in Q_j\}} + \beta_\tau^{(d)}\log(RV_t) \\ &+ \beta_\tau^{(w)}\log(\overline{RV}_{t-4,t}) + \beta_\tau^{(m)}\log(\overline{RV}_{t-21,t}) \\ &+ \gamma_\tau\log(RV_t)\mathbf{1}_{\{r_{sp,t}<0\}} + \beta_\tau^{(H)}\log(\overline{HCV}_{t-21,t}) + \omega_{t+\tau} \quad (4.28)\end{aligned}$$

With significantly positive values of estimated $\beta_\tau^{(H)}$, the lagged monthly cross-sectional housing return dispersion shows predictability for both housing and stock realized variances for all horizons from daily to monthly. In addition, for predicting housing realized variances, the inclusion of $\log(\overline{HCV}_{t-21,t})$ dramatically increases the R^2s for all horizons. In particular, for the one-month-ahead forecast of housing realized variance, the R^2 increases from 0.488 to 0.611. If we replace the housing cross-sectional variance with that of the stock market, the coefficient of $\overline{SCV}_{t-21,t}$ is significantly positive in (4.29) for only the monthly forecast horizon but is significantly positive at all horizons in (4.30), which implies that the stock cross-sectional variance is only correlated with housing volatility in the long-run, although the longer-run relationship with housing may come from the gap between deal closure and record dates.

$$\begin{aligned}\log(\overline{HRV}_{t+1,t+\tau}) &= c_\tau + \sum_{j=1}^{3}\beta_{j,\tau}\mathbf{1}_{\{t+\tau\in Q_j\}} + \beta_\tau^{(d)}\log(HRV_t) \\ &+ \beta_\tau^{(w)}\log(\overline{HRV}_{t-4,t}) + \beta_\tau^{(m)}\log(\overline{HRV}_{t-21,t}) \\ &+ \gamma_\tau\log(HRV_t)\mathbf{1}_{\{r_{c,t}<0\}} + \beta_\tau^{(S)}\log(\overline{SCV}_{t-21,t}) + \omega_{t+\tau} \quad (4.29)\end{aligned}$$

$$\begin{aligned}\log(\overline{RV}_{t+1,t+\tau}) &= c_\tau + \sum_{j=1}^{3}\beta_{j,\tau}\mathbf{1}_{\{t+\tau\in Q_j\}} + \beta_\tau^{(d)}\log(RV_t) \\ &+ \beta_\tau^{(w)}\log(\overline{RV}_{t-4,t}) + \beta_\tau^{(m)}\log(\overline{RV}_{t-21,t}) \\ &+ \gamma_\tau\log(RV_t)\mathbf{1}_{\{r_{sp,t}<0\}} + \beta_\tau^{(S)}\log(\overline{SCV}_{t-21,t}) + \omega_{t+\tau} \quad (4.30)\end{aligned}$$

If all three variance measures, realized variances, and two cross-sectional variances are included in the predictive regressions, as Table 4.7 illustrates, stock realized variance still shows predictability for housing realized variances, but this predictability is not as strong as that of the cross-sectional variance[15]. The stock realized variance and stock cross-sectional variance take opposite signs for all horizons, which may result from the high correlation between these two measures. Wald tests for joint significance of the three variables suggest that they jointly carry information that is useful in forecasting housing realized variances for all horizons. For predicting stock realized variance, only the coefficient of stock cross-sectional variance at the daily horizon is significant. However, the Wald tests suggest that $\log(\overline{HRV}_{t-21,t})$, $\log(\overline{HCV}_{t-21,t})$ and $\log(\overline{SCV}_{t-21,t})$ are still jointly significant at the conventional significance level for all forecast horizons, which implies that these three measures carry similar useful information on predicting stock realized variance. The R^2s for predicting housing realized variances increase dramatically with the inclusion of all three measures.

4.6.3 Out-of-sample forecast performance of cross-sectional variances

Overall, the in-sample regression results imply that the two cross-sectional variance measures are potentially the best predictors of realized variances. It is also instructive to evaluate the out-of-sample forecast performance of each variance measure. The benchmark model is the HAR-RV model with a volatility asymmetry term and quarterly dummies, whose forecast performance is compared with that of four other models, each of which is a benchmark model augmented with realized, GJR-GARCH variance of the other market or the two cross-sectional variances. The in-sample period is June 2001 to June 2009, and July 2009 to December 2011 is the out-of-sample

[15] The GJR-GARCH estimated variances are excluded, because they are highly correlated with the realized measures, which could cause the problem of collinearity.

period[16]. The model parameters are estimated once over the fixed in-sample window. Table 4.8 reports the Diebold-Mariano test statistics and corresponding P-values for equal forecast accuracy against the alternative that the forecasts from the augmented HAR-RV model outperforms the benchmark model under the MSE loss function. The Diebold-Mariano tests indicate that only the cross-sectional variance measures are useful for out-of-sample forecasts of realized variances. In particular, including cross-sectional housing variance leads to significantly better out-of-sample forecasts for housing realized variances than those from the benchmark for all horizons from daily to monthly. The augmented model with stock cross-sectional variance can beat the performance of the benchmark HAR-RV model for predicting realized stock variance only at the daily forecast horizon.

4.7 Conclusion

This paper uses a set of newly constructed daily housing price series, as well as daily and high-frequency intraday stock prices, to investigate the volatility asymmetries and volatility relationship of housing and stock markets. Results based on a set of predictive regressions suggest that significant volatility asymmetries exist in the aggregate housing market and seven out of ten individual areas. A direct test of volatility on changes in loan-to-value ratio suggests that the observed volatility asymmetry in the housing market does not come from changes in degree of housing financial leverage, but instead results from the risk premium carried by housing volatility, as illustrated by a consumption-based asset pricing model with housing. The volatilities of the two markets are found to be positively correlated; in particular, if the volatility in one market increases (decreases), higher (lower) future volatility in the other market will be observed. Moreover, the cross-sectional vari-

[16] Earlier version of the daily house price indexes is based on a vintage of the DataQuick database that ended in June 2009, which is how the sample-split point is chosen.

ances are positively correlated with the future volatilities of the two markets, and housing cross-sectional variance can be used to improve the out-of-sample forecast performance for housing realized variances.

The linkage between housing and stock volatilities could also be investigated in local areas where there are leading industries or firms, and local housing volatility might be closely linked to the stock volatility of local firms. In addition, besides the daily housing price indexes, other measures, such as a housing liquidity index, that concern the microstructure of housing market also could be constructed, which may be useful to predict housing and stock realized volatilities. Furthermore, the volatility dynamics of housing prices and its relationship to financial markets are fundamental to the valuations of many financial derivatives, such as housing index options and futures, mortgage insurance, and mortgage-backed securities. The more accurate volatility measures afforded by higher frequency housing indexes could facilitate more accurate pricing and forecasts of those financial derivatives. I leave these for future research.

Table 4.1: Unconditional correlations of HRV, RV, HCV and SCV

	daily				weekly				monthly			
	HRV	RV	HCV	SCV	HRV	RV	HCV	SCV	HRV	RV	HCV	SCV
HRV	1.000	0.082	0.241	0.083	1.000	0.189	0.481	0.180	1.000	0.307	0.700	0.316
RV		1.000	0.292	0.380		1.000	0.374	0.557		1.000	0.433	0.682
HCV			1.000	0.263			1.000	0.421			1.000	0.562
SCV				1.000				1.000				1.000

Note: HRV and RV are housing and stock realized variances, respectively. HCV and SCV are housing and stock cross-sectional variances, respectively. All three variance measures are based on data from June 2001 to December 2011.

Table 4.2: Summary Statistics

	Mean	Std. dev.	Median	Skewness	Exc. Kurt
		Panel A: HRV_t			
$HRV_t \times 100$	0.006	0.004	0.005	1.699	4.302
Annualized $\sqrt{HRV_t}$	2.625	0.800	2.528	0.730	0.705
$\log(HRV_t)$	-9.857	0.609	-9.840	-0.094	-0.279
		Panel B: RV_t			
$RV_t \times 100$	0.397	0.632	0.203	4.757	28.321
Annualized $\sqrt{RV_t}$	18.726	11.209	15.620	2.348	7.547
$\log(RV_t)$	-6.094	0.962	-6.198	0.730	0.180
		Panel C: HCV_t			
$HCV_t \times 100$	1.682	0.651	1.393	1.102	0.292
Annualized $\sqrt{HCV_t}$	12.754	2.357	11.802	0.863	-0.361
$\log(HCV_t)$	-4.150	0.350	-4.274	0.648	-0.803
		Panel D: SCV_t			
$SCV_t \times 100$	0.032	0.046	0.015	3.662	15.789
Annualized $\sqrt{SCV_t}$	1.566	0.868	1.243	2.319	5.790
$\log(SCV_t)$	-8.522	0.842	-8.775	1.199	1.004

Note: RV_t and HRV_t are monthly housing and stock realized variances at date t. HCV_t and SCV_t are the annualized monthly average housing and stock cross-sectional variances respectively. $\sqrt{RV_t}$, $\sqrt{HRV_t}$, $\sqrt{HCV_t}$, $\sqrt{SCV_t}$ are annualized monthly volatilities (Std.dev) in percentage that correspond to Figure 4.2.

Table 4.3: Leverage effect in predictive regressions for individual housing markets

$$\log(\overline{HRV}_{i,t+1,t+\tau}) =$$
$$c_{i,\tau} + \sum_{j=1}^{3} \beta_{i,j,\tau}\mathbf{1}_{\{t+\tau\in Q_j\}} + \beta_{i,\tau}^{(d)} \log(HRV_{i,t}) + \beta_{i,\tau}^{(w)} \log(\overline{HRV}_{i,t-4,t}) + \beta_{i,\tau}^{(m)} \log(\overline{HRV}_{i,t-21,t}) + \gamma_{i,\tau} \log(HRV_{i,t})\mathbf{1}_{\{r_{i,t}<0\}} + \omega_{i,t+\tau}$$

	Los Angeles			Boston			Chicago			Denver			Miami		
	τ=1	τ=5	τ=22	τ=1	τ=5	τ=22	τ=1	τ=5	τ=22	τ=1	τ=5	τ=22	τ=1	τ=5	τ=22
$c_{i,\tau}$	-4.252	-3.574	-3.613	-6.410	-5.405	-6.337	-2.652	-2.355	-1.955	-6.502	-5.883	-6.913	-2.941	-2.045	-2.037
	(0.767)	(0.527)	(0.498)	(1.012)	(0.700)	(0.820)	(0.539)	(0.422)	(0.465)	(0.898)	(0.702)	(0.834)	(0.554)	(0.423)	(0.481)
$\beta_{i,1,\tau}$	-0.194	-0.026	0.057	0.062	0.215	0.336	0.040	0.070	0.040	0.061	-0.015	0.155	0.039	-0.001	0.009
	(0.119)	(0.084)	(0.086)	(0.142)	(0.095)	(0.075)	(0.121)	(0.094)	(0.102)	(0.132)	(0.096)	(0.128)	(0.122)	(0.093)	(0.091)
$\beta_{i,2,\tau}$	-0.241	-0.076	-0.089	-0.262	-0.320	-0.161	-0.174	-0.116	-0.241	-0.117	-0.225	-0.271	0.015	-0.039	-0.115
	(0.113)	(0.079)	(0.085)	(0.126)	(0.096)	(0.089)	(0.131)	(0.092)	(0.111)	(0.118)	(0.099)	(0.118)	(0.115)	(0.083)	(0.080)
$\beta_{i,3,\tau}$	-0.291	-0.070	-0.061	-0.076	-0.183	-0.265	-0.129	-0.116	-0.202	-0.126	-0.206	-0.247	0.082	0.069	0.079
	(0.105)	(0.075)	(0.077)	(0.132)	(0.098)	(0.092)	(0.127)	(0.102)	(0.121)	(0.125)	(0.098)	(0.098)	(0.102)	(0.079)	(0.089)
$\beta_{i,\tau}^{(d)}$	-0.009	0.011	0.003	0.046	0.023	0.008	-0.035	0.016	0.005	0.026	0.016	0.002	0.006	-0.002	0.000
	(0.021)	(0.008)	(0.003)	(0.021)	(0.007)	(0.004)	(0.019)	(0.008)	(0.005)	(0.022)	(0.007)	(0.004)	(0.021)	(0.007)	(0.003)
$\beta_{i,\tau}^{(w)}$	0.104	-0.014	0.025	0.030	-0.045	0.035	0.158	0.041	0.064	0.053	-0.040	0.021	0.004	-0.016	0.021
	(0.062)	(0.039)	(0.018)	(0.058)	(0.038)	(0.019)	(0.075)	(0.057)	(0.042)	(0.061)	(0.032)	(0.020)	(0.069)	(0.041)	(0.019)
$\beta_{i,\tau}^{(m)}$	0.633	0.706	0.659	0.483	0.567	0.404	0.738	0.735	0.734	0.477	0.535	0.383	0.840	0.853	0.798
	(0.090)	(0.067)	(0.052)	(0.107)	(0.078)	(0.077)	(0.095)	(0.080)	(0.073)	(0.097)	(0.067)	(0.072)	(0.079)	(0.055)	(0.050)
$\gamma_{i,\tau}(\times 10^{-1})$	-0.111	-0.037	-0.071	-0.121	-0.047	-0.032	0.127	0.036	0.016	-0.044	-0.015	-0.025	-0.036	-0.015	-0.039
	(0.064)	(0.023)	(0.016)	(0.074)	(0.026)	(0.014)	(0.068)	(0.029)	(0.021)	(0.070)	(0.024)	(0.013)	(0.069)	(0.024)	(0.015)
R^2	0.043	0.234	0.501	0.029	0.206	0.425	0.124	0.447	0.647	0.027	0.144	0.308	0.087	0.424	0.667

$$\log(\overline{HRV}_{i,t+1,t+\tau}) =$$
$$c_{i,\tau} + \sum_{j=1}^{3} \beta_{i,j,\tau} \mathbf{1}_{\{t+\tau \in Q_j\}} + \beta_{i,\tau}^{(d)} \log(HRV_{i,t}) + \beta_{i,\tau}^{(w)} \log(\overline{HRV}_{i,t-4,t}) + \beta_{i,\tau}^{(m)} \log(\overline{HRV}_{i,t-21,t}) + \gamma_{i,\tau} \log(HRV_{i,t}) \mathbf{1}_{\{r_{i,t}<0\}} + \omega_{i,t+\tau}$$

	Las Vegas			San Diego			San Francisco			New York			Washington, D.C.		
	τ=1	τ=5	τ=22	τ=1	τ=5	τ=22	τ=1	τ=5	τ=22	τ=1	τ=5	τ=22	τ=1	τ=5	τ=22
$c_{i,\tau}$	-5.563	-4.038	-3.870	-7.189	-5.383	-5.742	-4.052	-2.437	-2.691	-5.408	-4.824	-5.339	-3.658	-3.267	-3.352
	(0.739)	(0.545)	(0.602)	(1.054)	(0.700)	(0.740)	(0.588)	(0.395)	(0.403)	(0.837)	(0.669)	(0.627)	(0.693)	(0.537)	(0.625)
$\beta_{i,1,\tau}$	0.102	0.021	0.127	-0.085	-0.071	0.067	-0.169	0.003	-0.056	-0.101	-0.018	0.082	-0.014	-0.102	-0.006
	(0.111)	(0.091)	(0.084)	(0.127)	(0.097)	(0.102)	(0.136)	(0.103)	(0.108)	(0.117)	(0.092)	(0.093)	(0.125)	(0.090)	(0.102)
$\beta_{i,2,\tau}$	0.106	-0.033	-0.075	-0.168	-0.180	-0.092	-0.032	-0.024	-0.237	-0.190	-0.192	-0.183	-0.078	-0.151	-0.169
	(0.103)	(0.090)	(0.107)	(0.135)	(0.088)	(0.084)	(0.127)	(0.093)	(0.098)	(0.109)	(0.092)	(0.099)	(0.127)	(0.100)	(0.113)
$\beta_{i,3,\tau}$	-0.124	-0.052	-0.004	-0.190	-0.179	-0.173	-0.056	0.070	-0.101	-0.134	-0.170	-0.263	0.010	-0.098	-0.202
	(0.121)	(0.092)	(0.111)	(0.124)	(0.081)	(0.070)	(0.126)	(0.099)	(0.097)	(0.116)	(0.083)	(0.081)	(0.129)	(0.095)	(0.106)
$\beta_{i,\tau}^{(d)}$	-0.041	0.007	0.003	0.001	0.015	0.003	-0.038	0.003	-0.004	-0.008	0.004	0.004	0.020	0.017	-0.001
	(0.022)	(0.006)	(0.004)	(0.022)	(0.006)	(0.003)	(0.021)	(0.007)	(0.004)	(0.019)	(0.007)	(0.004)	(0.022)	(0.007)	(0.003)
$\beta_{i,\tau}^{(w)}$	0.128	0.010	0.044	0.054	-0.033	0.009	0.072	0.039	0.080	0.034	0.029	0.022	0.211	0.095	0.018
	(0.061)	(0.039)	(0.022)	(0.064)	(0.039)	(0.018)	(0.065)	(0.040)	(0.026)	(0.056)	(0.039)	(0.018)	(0.072)	(0.044)	(0.018)
$\beta_{i,\tau}^{(m)}$	0.549	0.647	0.618	0.412	0.545	0.474	0.711	0.754	0.666	0.618	0.572	0.517	0.536	0.594	0.665
	(0.082)	(0.064)	(0.058)	(0.103)	(0.078)	(0.076)	(0.081)	(0.051)	(0.052)	(0.090)	(0.063)	(0.059)	(0.089)	(0.064)	(0.062)
$\gamma_{i,\tau}(\times 10^{-1})$	-0.026	0.004	-0.038	-0.009	0.037	-0.045	0.060	-0.015	0.007	-0.034	-0.065	-0.033	0.096	0.004	-0.001
	(0.068)	(0.024)	(0.019)	(0.069)	(0.027)	(0.014)	(0.070)	(0.025)	(0.014)	(0.060)	(0.025)	(0.014)	(0.072)	(0.027)	(0.015)
R^2	0.044	0.239	0.465	0.013	0.116	0.311	0.068	0.371	0.578	0.030	0.181	0.403	0.061	0.273	0.510

Note: Newey-West standard errors with lags of $\lfloor T^{1/3} \rfloor + 2\tau$ are shown in parentheses.

Table 4.4: Leverage effect in predictive regressions with changes in LTV

$$\log(\overline{HRV}_{i,t+1,t+\tau}) = c_{i,\tau} + \sum_{j=1}^{3} \beta_{i,j,\tau}\mathbf{1}_{\{t+\tau\in Q_j\}}$$
$$+\beta_{i,\tau}^{(d)}\log(HRV_{i,t}) + \beta_{i,\tau}^{(w)}\log(\overline{HRV}_{i,t-4,t}) + \beta_{i,\tau}^{(m)}\log(\overline{HRV}_{i,t-21,t}) + (\gamma_{i,\tau} + \gamma_{i,\tau}^{'}\triangle LTV_{i,t})\log(HRV_{i,t})\mathbf{1}_{\{r_{i,t}<0\}} + \omega_{i,t+\tau}$$

	Los Angeles			Boston			Chicago			Denver			Miami		
	τ=1	τ=5	τ=22	τ=1	τ=5	τ=22	τ=1	τ=5	τ=22	τ=1	τ=5	τ=22	τ=1	τ=5	τ=22
$c_{i,\tau}$	-4.249	-3.575	-3.614	-6.402	-5.399	-6.334	-2.705	-2.362	-1.966	-6.532	-5.909	-6.923	-2.954	-2.050	-2.037
	(0.767)	(0.527)	(0.497)	(1.011)	(0.699)	(0.820)	(0.541)	(0.422)	(0.466)	(0.902)	(0.701)	(0.835)	(0.555)	(0.423)	(0.480)
$\beta_{i,1,\tau}$	-0.194	-0.025	0.057	0.060	0.213	0.335	0.037	0.070	0.039	0.063	-0.014	0.155	0.039	-0.001	0.009
	(0.119)	(0.083)	(0.086)	(0.141)	(0.095)	(0.075)	(0.121)	(0.094)	(0.102)	(0.132)	(0.096)	(0.128)	(0.122)	(0.093)	(0.091)
$\beta_{i,2,\tau}$	-0.241	-0.076	-0.088	-0.264	-0.321	-0.161	-0.174	-0.116	-0.242	-0.118	-0.225	-0.270	0.015	-0.038	-0.115
	(0.113)	(0.079)	(0.085)	(0.126)	(0.096)	(0.089)	(0.131)	(0.092)	(0.111)	(0.118)	(0.099)	(0.118)	(0.115)	(0.083)	(0.080)
$\beta_{i,3,\tau}$	-0.291	-0.070	-0.061	-0.077	-0.183	-0.265	-0.132	-0.116	-0.203	-0.127	-0.207	-0.247	0.083	0.069	0.079
	(0.105)	(0.075)	(0.077)	(0.132)	(0.098)	(0.092)	(0.127)	(0.102)	(0.120)	(0.125)	(0.097)	(0.098)	(0.102)	(0.079)	(0.089)
$\beta_{i,\tau}^{(d)}$	-0.009	0.011	0.003	0.046	0.023	0.008	-0.037	0.016	0.004	0.026	0.016	0.002	0.005	-0.002	0.000
	(0.021)	(0.008)	(0.003)	(0.021)	(0.007)	(0.004)	(0.019)	(0.008)	(0.005)	(0.022)	(0.007)	(0.004)	(0.021)	(0.007)	(0.003)
$\beta_{i,\tau}^{(w)}$	0.104	-0.014	0.025	0.030	-0.045	0.035	0.158	0.041	0.064	0.053	-0.040	0.021	0.005	-0.016	0.021
	(0.062)	(0.039)	(0.018)	(0.058)	(0.038)	(0.019)	(0.074)	(0.057)	(0.042)	(0.061)	(0.032)	(0.020)	(0.069)	(0.041)	(0.019)
$\beta_{i,\tau}^{(m)}$	0.634	0.705	0.658	0.484	0.568	0.404	0.736	0.735	0.734	0.475	0.533	0.382	0.839	0.853	0.798
	(0.090)	(0.067)	(0.052)	(0.107)	(0.077)	(0.077)	(0.095)	(0.080)	(0.073)	(0.098)	(0.067)	(0.072)	(0.079)	(0.055)	(0.050)
$\gamma_{i,\tau}(\times 10^{-1})$	-0.110	-0.037	-0.072	-0.121	-0.047	-0.032	0.133	0.037	0.017	-0.042	-0.013	-0.025	-0.035	-0.015	-0.039
	(0.064)	(0.023)	(0.016)	(0.074)	(0.026)	(0.014)	(0.068)	(0.029)	(0.022)	(0.070)	(0.024)	(0.013)	(0.069)	(0.024)	(0.015)
$\gamma_{i,\tau}^{'}$	0.086	-0.027	-0.034	0.057	0.039	0.022	0.206	0.029	0.047	-0.057	-0.046	-0.018	0.065	0.026	-0.003
	(0.152)	(0.047)	(0.023)	(0.104)	(0.032)	(0.016)	(0.095)	(0.033)	(0.021)	(0.086)	(0.025)	(0.016)	(0.100)	(0.032)	(0.015)
R^2	0.043	0.234	0.501	0.029	0.206	0.426	0.125	0.447	0.647	0.027	0.145	0.308	0.088	0.424	0.667

$$\log(\overline{HRV}_{i,t+1,t+\tau}) = c_{i,\tau} + \sum_{j=1}^{3} \beta_{i,j,\tau}\mathbf{1}_{\{t+\tau \in Q_j\}}$$
$$+\beta_{i,\tau}^{(d)} \log(HRV_{i,t}) + \beta_{i,\tau}^{(w)} \log(\overline{HRV}_{i,t-4,t}) + \beta_{i,\tau}^{(m)} \log(\overline{HRV}_{i,t-21,t}) + (\gamma_{i,\tau} + \gamma_{i,\tau}^{'} \triangle LTV_{i,t}) \log(HRV_{i,t})\mathbf{1}_{\{r_{i,t}<0\}} + \omega_{i,t+\tau}$$

	Las Vegas			San Diego			San Francisco			New York			Washington, D.C.		
	τ=1	τ=5	τ=22	τ=1	τ=5	τ=22	τ=1	τ=5	τ=22	τ=1	τ=5	τ=22	τ=1	τ=5	τ=22
$c_{i,\tau}$	-5.564	-4.038	-3.868	-7.181	-5.389	-5.744	-4.011	-2.439	-2.691	-5.402	-4.822	-5.339	-3.614	-3.263	-3.348
	(0.739)	(0.545)	(0.600)	(1.054)	(0.701)	(0.740)	(0.588)	(0.396)	(0.404)	(0.834)	(0.669)	(0.627)	(0.688)	(0.536)	(0.625)
$\beta_{i,1,\tau}$	0.101	0.022	0.126	-0.084	-0.071	0.067	-0.168	0.003	-0.056	-0.104	-0.019	0.082	-0.015	-0.102	-0.007
	(0.111)	(0.091)	(0.084)	(0.127)	(0.097)	(0.102)	(0.136)	(0.103)	(0.108)	(0.116)	(0.092)	(0.094)	(0.124)	(0.090)	(0.102)
$\beta_{i,2,\tau}$	0.106	-0.033	-0.075	-0.167	-0.181	-0.092	-0.027	-0.024	-0.237	-0.190	-0.192	-0.184	-0.079	-0.151	-0.170
	(0.103)	(0.090)	(0.107)	(0.135)	(0.087)	(0.084)	(0.127)	(0.093)	(0.098)	(0.108)	(0.092)	(0.098)	(0.126)	(0.100)	(0.113)
$\beta_{i,3,\tau}$	-0.125	-0.052	-0.004	-0.188	-0.180	-0.174	-0.052	0.070	-0.101	-0.137	-0.172	-0.264	0.009	-0.099	-0.202
	(0.121)	(0.092)	(0.111)	(0.124)	(0.081)	(0.070)	(0.126)	(0.098)	(0.097)	(0.116)	(0.083)	(0.080)	(0.129)	(0.095)	(0.106)
$\beta_{i,\tau}^{(d)}$	-0.041	0.007	0.004	0.000	0.015	0.003	-0.035	0.002	-0.004	-0.007	0.005	0.004	0.024	0.018	-0.001
	(0.022)	(0.006)	(0.004)	(0.022)	(0.006)	(0.003)	(0.021)	(0.007)	(0.004)	(0.019)	(0.007)	(0.004)	(0.022)	(0.007)	(0.003)
$\beta_{i,\tau}^{(w)}$	0.128	0.010	0.043	0.054	-0.032	0.009	0.070	0.039	0.080	0.033	0.028	0.022	0.209	0.095	0.018
	(0.061)	(0.039)	(0.022)	(0.064)	(0.039)	(0.018)	(0.065)	(0.040)	(0.026)	(0.056)	(0.039)	(0.018)	(0.072)	(0.044)	(0.018)
$\beta_{i,\tau}^{(m)}$	0.549	0.647	0.618	0.414	0.544	0.474	0.714	0.754	0.666	0.619	0.572	0.517	0.538	0.594	0.665
	(0.082)	(0.064)	(0.058)	(0.103)	(0.078)	(0.076)	(0.081)	(0.051)	(0.052)	(0.090)	(0.063)	(0.059)	(0.088)	(0.064)	(0.062)
$\gamma_{i,\tau}(\times 10^{-1})$	-0.026	0.004	-0.039	-0.009	0.037	-0.045	0.053	-0.014	0.008	-0.035	-0.065	-0.033	0.097	0.004	-0.001
	(0.068)	(0.024)	(0.019)	(0.069)	(0.027)	(0.014)	(0.070)	(0.025)	(0.014)	(0.060)	(0.025)	(0.014)	(0.072)	(0.027)	(0.015)
$\gamma_{i,\tau}^{'}$	0.034	-0.006	-0.035	0.048	-0.030	-0.015	-0.151	0.008	0.003	0.132	0.058	0.030	-0.405	-0.046	-0.030
	(0.104)	(0.034)	(0.014)	(0.075)	(0.026)	(0.014)	(0.101)	(0.032)	(0.019)	(0.118)	(0.034)	(0.021)	(0.115)	(0.038)	(0.017)
R^2	0.044	0.239	0.465	0.014	0.116	0.311	0.068	0.371	0.578	0.031	0.182	0.403	0.065	0.273	0.511

Note: Newey-West standard errors with lags of $\lfloor T^{1/3} \rfloor + 2\tau$ are shown in parentheses.

Table 4.5: GJR-GARCH model estimation results for individual areas

$$r_{i,t} = \mu_i + \rho_{i1} r_{i,t-1} + \rho_{i5} r_{i,t-5} + \rho_{im} r^m_{i,t-1} + \rho_{ic} r^m_{c,t-1} + \varepsilon_{i,t}$$
$$Var_{t-1}(r_{i,t}) \equiv h_{i,t} = \omega_i + \alpha_i \varepsilon^2_{i,t-1} + \beta_i h_{i,t-1} + \gamma_i \varepsilon^2_{i,t-1} \mathbf{1}_{\{\varepsilon_{i,t-1}<0\}}$$

	Los Angeles	Boston	Chicago	Denver	Miami	Las Vegas	San Diego	San Francisco	New York	D.C.
				Panel A: Conditional Mean						
μ_i	0.015	-0.003	0.000	-0.002	0.013	-0.003	-0.001	-0.003	0.009	0.010
	(0.007)	(0.007)	(0.010)	(0.006)	(0.007)	(0.007)	(0.007)	(0.010)	(0.005)	(0.009)
$r_{i,t-1}$	-0.082	0.024	0.014	-0.008	-0.029	-0.003	-0.044	-0.096	0.044	0.011
	(0.021)	(0.023)	(0.022)	(0.023)	(0.020)	(0.022)	(0.020)	(0.019)	(0.021)	(0.020)
$r_{i,t-5}$	0.052	0.003	-0.004	0.005	-0.015	0.003	-0.033	0.155	0.005	0.029
	(0.020)	(0.021)	(0.021)	(0.023)	(0.020)	(0.023)	(0.021)	(0.024)	(0.021)	(0.021)
$r^m_{i,t-1}$	-0.009	-0.011	-0.027	-0.008	-0.009	0.018	-0.012	-0.013	-0.030	-0.032
	(0.007)	(0.005)	(0.007)	(0.005)	(0.006)	(0.004)	(0.006)	(0.006)	(0.006)	(0.007)
$r^m_{c,t-1}$	0.049	0.031	0.041	0.014	0.050	0.028	0.056	0.048	0.054	0.080
	(0.009)	(0.006)	(0.009)	(0.005)	(0.009)	(0.008)	(0.008)	(0.010)	(0.006)	(0.010)
				Panel B: Conditional Variance						
$\omega_i(\times 10^{-2})$	0.032	0.221	0.077	0.226	0.018	0.019	0.024	0.033	0.036	0.063
	(0.018)	(0.076)	(0.056)	(0.169)	(0.018)	(0.014)	(0.025)	(0.029)	(0.021)	(0.040)
$\varepsilon^2_{i,t-1}$	0.016	0.044	0.039	0.032	0.003	0.014	0.013	0.010	0.031	0.036
	(0.006)	(0.011)	(0.013)	(0.021)	(0.009)	(0.007)	(0.006)	(0.006)	(0.008)	(0.008)
$h_{i,t-1}$	0.974	0.930	0.948	0.943	0.988	0.981	0.984	0.984	0.970	0.966
	(0.006)	(0.012)	(0.009)	(0.029)	(0.004)	(0.006)	(0.005)	(0.005)	(0.006)	(0.006)
$\varepsilon^2_{i,t-1}\mathbf{1}_{\{\varepsilon_{i,t-1}<0\}}$	0.014	0.016	0.030	0.004	0.016	0.008	0.004	0.009	-0.009	-0.008
	(0.009)	(0.015)	(0.023)	(0.021)	(0.010)	(0.009)	(0.006)	(0.006)	(0.011)	(0.013)

Note: This table shows Quasi Maximum Likelihood Estimates (QMLE) estimates with robust standard errors in parentheses.

Table 4.6: GJR-GARCH model estimation results for aggregate indexes

$$r_{c,t} = \mu_c + \sum_{j=1}^{5} \rho_j \ r_{c,t-j} + \rho_m \ r^m_{c,t-1} + \varepsilon_{c,t}$$
$$Var_{t-1}(r_{c,t}) \equiv h_{c,t} = \omega + \alpha \ \varepsilon^2_{c,t-1} + \beta \ h_{c,t-1} + \gamma \ \varepsilon^2_{c,t-1}\mathbf{1}_{\{\varepsilon_{c,t-1}<0\}}$$

$$r_{sp,t} = \mu_{sp} + \rho_1 \ r_{sp,t-1} + \varepsilon_{sp,t}$$
$$Var_{t-1}(r_{sp,t}) \equiv h_{sp,t} = \omega + \alpha \ \varepsilon^2_{sp,t-1} + \beta \ h_{sp,t-1} + \gamma \ \varepsilon^2_{sp,t-1}\mathbf{1}_{\{\varepsilon_{sp,t-1}<0\}}$$

	$r_{c,t}$		$r_{sp,t}$	
	Panel A: Conditional mean			
μ	0.004	0.003	0.045	0.006
	(0.003)	(0.003)	(0.017)	(0.018)
r_{t-1}	-0.083	-0.078	-0.074	-0.066
	(0.020)	(0.020)	(0.019)	(0.018)
r_{t-2}	0.018	0.021		
	(0.021)	(0.021)		
r_{t-3}	-0.005	-0.007		
	(0.021)	(0.021)		
r_{t-4}	-0.016	-0.016		
	(0.022)	(0.022)		
r_{t-5}	0.061	0.057		
	(0.021)	(0.022)		
r^m_{t-1}	0.038	0.036		
	(0.003)	(0.003)		
	Panel B: Conditional variance			
$\omega(\times 10^{-2})$	0.006	0.006	1.340	1.222
	(0.004)	(0.002)	(0.579)	(0.333)
ε^2_{t-1}	0.030	-0.008	0.083	-0.022
	(0.006)	(0.004)	(0.014)	(0.010)
h_{t-1}	0.968	0.988	0.909	0.941
	(0.007)	(0.004)	(0.013)	(0.008)
$\varepsilon^2_{t-1}\mathbf{1}_{\{\varepsilon_{t-1}<0\}}$		0.036		0.137
		(0.009)		(0.017)

Note: This table shows Quasi Maximum Likelihood Estimates (QMLE) estimates with robust standard errors in parentheses.

Table 4.7: Estimation results of predictive regressions

Benchmark model: $\log(\overline{HRV}_{t+1,t+\tau}) =$

$$c_\tau + \sum_{j=1}^{3} \beta_{j,\tau} \mathbf{1}_{\{t+\tau \in Q_j\}} + \beta_\tau^{(d)} \log(HRV_t) + \beta_\tau^{(w)} \log(\overline{HRV}_{t-4,t}) + \beta_\tau^{(m)} \log(\overline{HRV}_{t-21,t}) + \gamma_\tau \log(HRV_t) \mathbf{1}_{\{r_{c,t}<0\}} + \omega_{t+\tau}$$

	τ=1						τ=5						τ=22					
	Panel A: $\log(\overline{HRV}_{t+1,t+\tau})$																	
c_τ	-5.165	-5.793	-5.503	-6.529	-5.065	-7.581	-3.565	-4.769	-3.979	-4.884	-3.388	-5.725	-4.445	-5.952	-4.865	-5.831	-4.169	-6.473
	(0.874)	(1.015)	(0.879)	(0.947)	(0.872)	(1.051)	(0.619)	(0.730)	(0.622)	(0.664)	(0.595)	(0.734)	(0.637)	(0.680)	(0.584)	(0.723)	(0.597)	(0.751)
$\beta_\tau^{(d)}$	-0.024	-0.024	-0.024	-0.026	-0.024	-0.026	0.002	0.002	0.002	0.000	0.002	-0.001	-0.004	-0.004	-0.004	-0.007	-0.004	-0.007
	(0.025)	(0.025)	(0.026)	(0.025)	(0.025)	(0.025)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.007)	(0.003)	(0.003)	(0.003)	(0.002)	(0.003)	(0.003)
$\beta_\tau^{(w)}$	0.053	0.052	0.051	0.050	0.053	0.048	0.008	0.007	0.004	0.004	0.007	0.001	0.028	0.026	0.022	0.022	0.026	0.018
	(0.068)	(0.068)	(0.068)	(0.068)	(0.068)	(0.068)	(0.042)	(0.042)	(0.041)	(0.042)	(0.042)	(0.041)	(0.022)	(0.021)	(0.020)	(0.019)	(0.021)	(0.018)
$\beta_\tau^{(m)}$	0.665	0.617	0.554	0.271	0.645	0.174	0.725	0.635	0.598	0.350	0.689	0.261	0.634	0.523	0.494	0.232	0.583	0.153
	(0.090)	(0.098)	(0.102)	(0.121)	(0.092)	(0.130)	(0.064)	(0.070)	(0.073)	(0.080)	(0.066)	(0.087)	(0.055)	(0.055)	(0.057)	(0.076)	(0.052)	(0.076)
γ_τ	-0.004	-0.004	-0.004	-0.002	-0.004	-0.001	-0.003	-0.003	-0.003	-0.001	-0.003	0.000	-0.005	-0.003	-0.004	-0.002	-0.004	-0.001
	(0.006)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
$\log(\overline{SV}^{GARCH}_{t+1,t+\tau})$		0.062						0.119						0.162				
		(0.051)						(0.038)						(0.041)				
$\log(\overline{RV}_{t-21,t})$			0.124			0.179			0.144			0.172			0.165			0.166
			(0.054)			(0.083)			(0.039)			(0.059)			(0.038)			(0.059)
$\log(\overline{HCV}_{t-21,t})$				0.912		0.981				0.867		0.876				0.940		0.910
				(0.191)		(0.182)				(0.125)		(0.124)				(0.117)		(0.121)
$\log(\overline{SCV}_{t-21,t})$					0.043	-0.196					0.077	-0.144					0.115	-0.110
					(0.055)	(0.084)					(0.040)	(0.060)					(0.037)	(0.056)
W						33.040						55.069						65.623
Pval						(0.000)						(0.000)						(0.000)
R^2	0.037	0.038	0.039	0.046	0.038	0.048	0.257	0.268	0.273	0.305	0.262	0.316	0.488	0.523	0.534	0.611	0.510	0.632

Benchmark model: $\log(\overline{RV}_{t+1,t+\tau}) =$

$$c_\tau + \sum_{j=1}^{3} \beta_{j,\tau} \mathbf{1}_{\{t+\tau \in Q_j\}} + \beta_\tau^{(d)} \log(RV_t) + \beta_\tau^{(w)} \log(\overline{RV}_{t-4,t}) + \beta_\tau^{(m)} \log(\overline{RV}_{t-21,t}) + \gamma_\tau \log(RV_t) \mathbf{1}_{\{r_{sp,t}<0\}} + \omega_{t+\tau}$$

	τ=1						τ=5						τ=22					
	Panel B: $\log(\overline{RV}_{t+1,t+\tau})$																	
c_τ	-0.766	-0.734	-0.562	-0.572	-0.516	-0.134	-0.844	-0.732	-0.127	-0.431	-0.599	0.251	-1.537	-1.203	0.707	-0.454	-1.130	1.308
	(0.125)	(0.128)	(0.311)	(0.155)	(0.131)	(0.320)	(0.187)	(0.192)	(0.516)	(0.248)	(0.216)	(0.553)	(0.370)	(0.390)	(1.089)	(0.580)	(0.453)	(1.210)
$\beta_\tau^{(d)}$	0.200	0.200	0.200	0.199	0.191	0.189	0.213	0.213	0.212	0.211	0.204	0.202	0.164	0.162	0.158	0.156	0.149	0.139
	(0.032)	(0.033)	(0.032)	(0.032)	(0.032)	(0.032)	(0.023)	(0.022)	(0.022)	(0.022)	(0.022)	(0.022)	(0.019)	(0.018)	(0.019)	(0.017)	(0.018)	(0.017)
$\beta_\tau^{(w)}$	0.519	0.520	0.521	0.519	0.508	0.510	0.462	0.461	0.467	0.462	0.451	0.454	0.366	0.360	0.379	0.366	0.345	0.352
	(0.039)	(0.040)	(0.040)	(0.040)	(0.039)	(0.040)	(0.045)	(0.044)	(0.045)	(0.045)	(0.045)	(0.045)	(0.063)	(0.063)	(0.063)	(0.061)	(0.058)	(0.058)
$\beta_\tau^{(m)}$	0.225	0.211	0.216	0.206	0.172	0.149	0.247	0.204	0.217	0.209	0.194	0.143	0.298	0.194	0.215	0.205	0.207	0.078
	(0.032)	(0.035)	(0.037)	(0.035)	(0.033)	(0.038)	(0.046)	(0.050)	(0.052)	(0.050)	(0.050)	(0.057)	(0.065)	(0.076)	(0.076)	(0.075)	(0.086)	(0.096)
γ_τ	-0.023	-0.023	-0.023	-0.023	-0.022	-0.023	-0.014	-0.014	-0.014	-0.014	-0.014	-0.014	-0.008	-0.009	-0.008	-0.009	-0.007	-0.008
	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
$\log(\overline{HV}^{GARCH}_{t+1,t+\tau})$		0.039						0.131						0.352				
		(0.028)						(0.047)						(0.103)				
$\log(\overline{HRV}_{t-21,t})$			0.020			0.029			0.072			0.060			0.224			0.182
			(0.029)			(0.034)			(0.047)			(0.056)			(0.090)			(0.111)
$\log(\overline{HCV}_{t-21,t})$				0.086		0.036				0.184		0.099				0.475		0.240
				(0.041)		(0.051)				(0.067)		(0.079)				(0.163)		(0.182)
$\log(\overline{SCV}_{t-21,t})$					0.109	0.115					0.108	0.119					0.185	0.215
					(0.023)	(0.024)					(0.038)	(0.040)					(0.090)	(0.097)
W						26.696						13.836						11.235
Pval						(0.000)						(0.003)						(0.011)
R^2	0.731	0.732	0.731	0.732	0.734	0.735	0.799	0.802	0.800	0.801	0.802	0.805	0.701	0.724	0.714	0.719	0.711	0.734

Note: This table shows the results of benchmark HAR-RV model augmented with different variance measures. Results of Wald test for joint significance of the augmented measures are also included. Newey-West standard errors with lags of $\lfloor T^{1/3} \rfloor + 2\tau$ are shown in parentheses.

Table 4.8: Diebold-Mariano forecast comparison tests

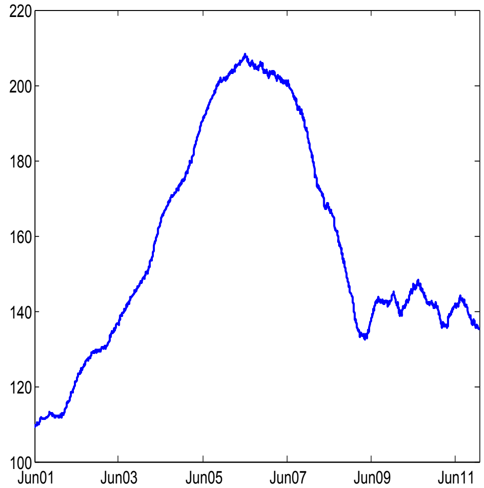
Panel A: $\log(\overline{HRV}_{t+1,t+\tau})$

τ=1				τ=5				τ=22			
$\overline{RV}_{t-21,t}$	$\overline{SV}^{GARCH}_{t+1,t+\tau}$	$\overline{HCV}_{t-21,t}$	$\overline{SCV}_{t-21,t}$	$\overline{RV}_{t-21,t}$	$\overline{SV}^{GARCH}_{t+1,t+\tau}$	$\overline{HCV}_{t-21,t}$	$\overline{SCV}_{t-21,t}$	$\overline{RV}_{t-21,t}$	$\overline{SV}^{GARCH}_{t+1,t+\tau}$	$\overline{HCV}_{t-21,t}$	$\overline{SCV}_{t-21,t}$
-0.963	-1.131	2.963	-3.087	-0.661	-0.788	1.701	-3.604	-4.238	-2.136	2.853	-5.205
(0.832)	(0.871)	(0.002)	(0.999)	(0.745)	(0.784)	(0.045)	(1.000)	(1.000)	(0.983)	(0.002)	(1.000)

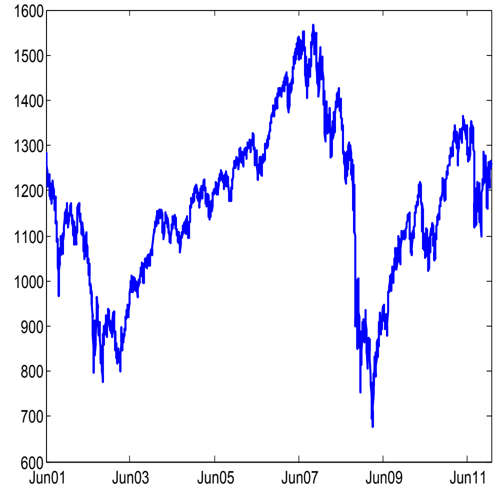
Panel B: $\log(\overline{RV}_{t+1,t+\tau})$

τ=1				τ=5				τ=22			
$\overline{HRV}_{t-21,t}$	$\overline{HV}^{GARCH}_{t+1,t+\tau}$	$\overline{HCV}_{t-21,t}$	$\overline{SCV}_{t-21,t}$	$\overline{HRV}_{t-21,t}$	$\overline{HV}^{GARCH}_{t+1,t+\tau}$	$\overline{HCV}_{t-21,t}$	$\overline{SCV}_{t-21,t}$	$\overline{HRV}_{t-21,t}$	$\overline{HV}^{GARCH}_{t+1,t+\tau}$	$\overline{HCV}_{t-21,t}$	$\overline{SCV}_{t-21,t}$
-0.975	-1.893	-1.369	1.526	0.199	-0.457	-0.214	-0.303	0.373	-0.816	-0.348	-1.025
(0.835)	(0.971)	(0.914)	(0.064)	(0.421)	(0.676)	(0.585)	(0.619)	(0.355)	(0.793)	(0.636)	(0.847)

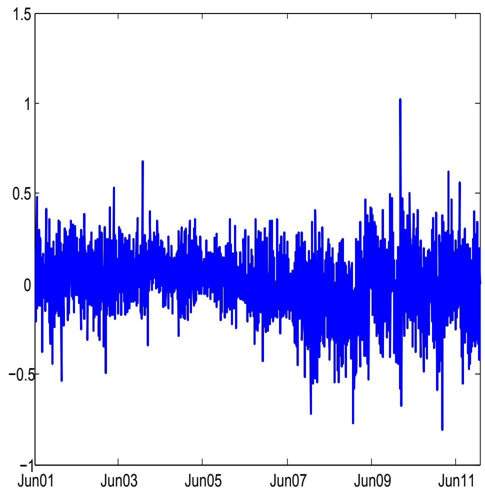
Note: The table reports the Diebold-Mariano tests for equal forecast accuracy against the alternative that the forecasts from HAR-RV augmented by realized variance, GJR-GARCH estimated variance or cross-sectional variance outperform the basic HAR-RV model under MSE loss function. Diebold-Mariano test statistics for forecasting housing and stock realized variances are respectively shown in Panel A and B, and the corresponding asymptotic p-values are reported in parentheses.



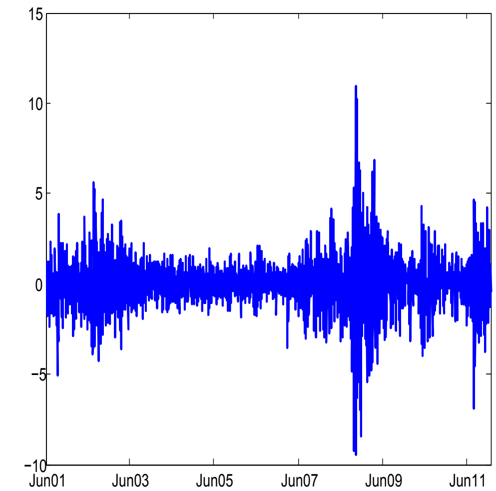
(a) Composite 10 index



(b) S&P 500 index

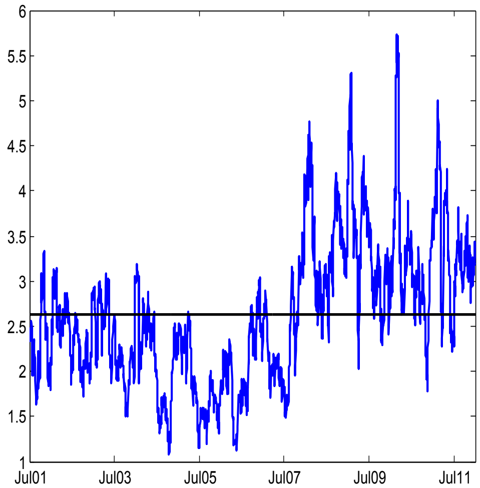


(c) Return of composite 10 index

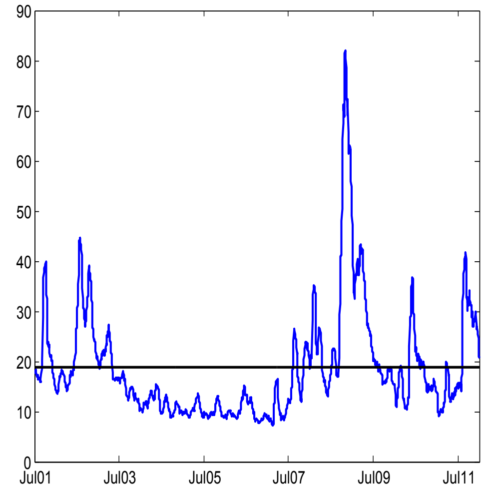


(d) Return of S&P 500 index

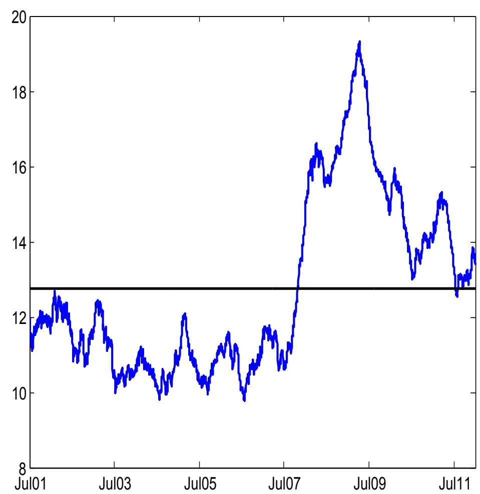
FIGURE 4.1: The daily indexes and returns of housing and stock markets



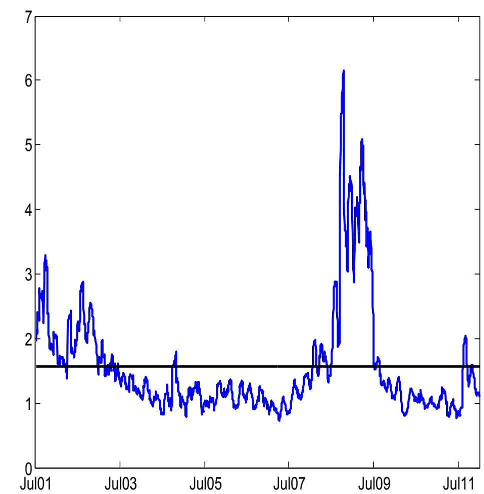
(a) Realized volatility of housing market



(b) Realized volatility of stock market



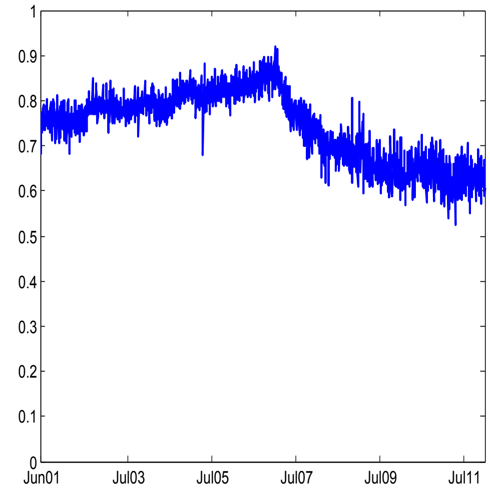
(c) Cross-sectional dispersion of housing market



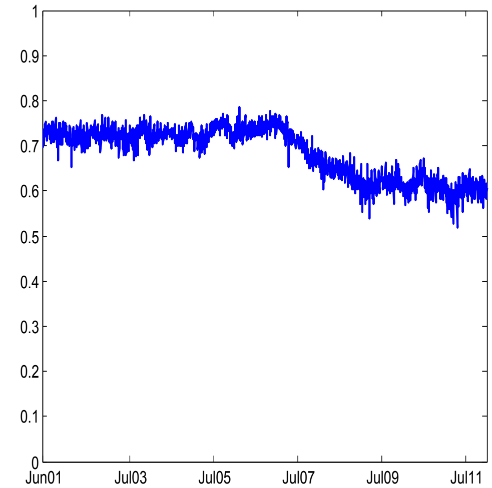
(d) Cross-sectional dispersion of stock market

Note: The realized volatilities for housing and stock market are annualized monthly realized volatility in percentage. The cross-sectional dispersions are the annualized monthly average volatilities in percentage.

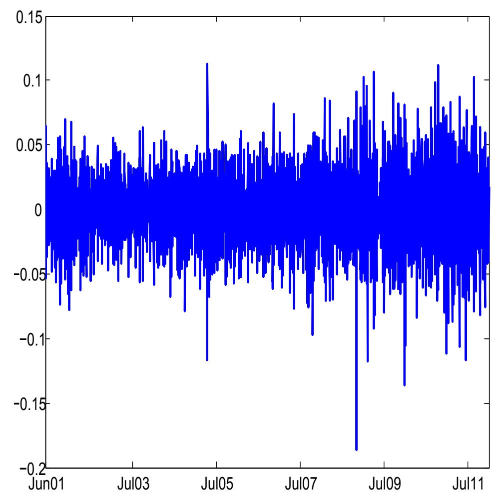
FIGURE 4.2: Realized volatilities and cross-sectional dispersions of housing and stock markets



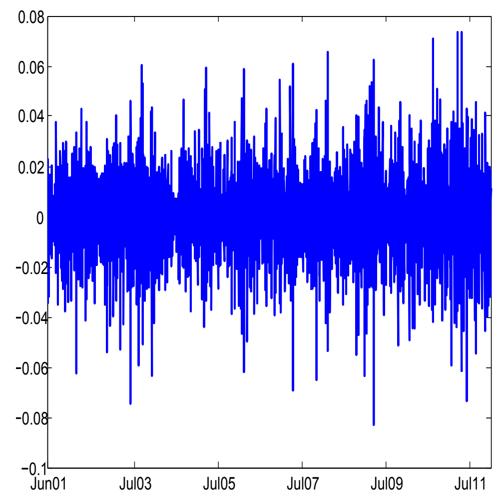
(a) Daily average LTV in Los Angeles



(b) Daily composite LTV



(c) Changes in daily LTV in Los Angeles



(d) Changes in daily composite LTV

FIGURE 4.3: Loan-to-value (LTV) ratios and changes in LTV

Appendix A

Appendix to Chapter 3

Table A.1: Metropolitan Statistical Areas (MSAs)

MSA	Represented counties	Counties in our indexes
Los Angeles-Long Beach-Santa Ana, CA Metropolitan Statistical Area (Los Angeles)	Los Angeles CA, Orange CA	Los Angeles CA, Orange CA
Boston-Cambridge-Quincy, MA-NH Metropolitan Statistical Area (Boston)	Essex MA, Middlesex MA, Norfolk MA, Plymouth MA, Suffolk MA, Rockingham NH, Strafford NH	Essex MA, Middlesex MA, Norfolk MA, Plymouth MA, Suffolk MA, Rockingham NH, Strafford NH
Chicago-Naperville-Joliet, IL Metropolitan Division (Chicago)	Cook IL, DeKalb IL, Du Page Il, Kane IL, Kendal IL, McHenry IL, Will IL, Grundy IL	Cook IL, DeKalb IL, Du Page Il, Kane IL, Kendal IL, McHenry IL, Will IL, Grundy IL
Denver-Aurora, CO Metropolitan Statistical Area (Denver)	Adams CO, Arapahoe CO, Broomfield CO, Clear Creek CO, Denver CO, Douglas CO, Elbert CO, Gilpin CO, Jefferson CO, Park CO	Adams CO, Arapahoe CO, Broomfield CO, Clear Creek CO, Denver CO, Douglas CO, Elbert CO, Gilpin CO, Jefferson CO, Park CO
Miami-Fort Lauderdale-Pompanp Beach, FL Metropolitan Statistical Area (Miami)	Broward FL, Miami-Dade FL, Palm Beach FL	Broward FL, Miami-Dade FL, Palm Beach FL
Las Vegas-Paradise, NV Metropolitan Statistical Area (Las Vegas)	Clark NV	Clark NV
San Diego-Carlsbad-San Marcos, CA Metropolitan Statistical Area (San Diego)	San Diego CA	San Diego CA
San Francisco-Oakland-Fremont, CA Metropolitan Statistical Area (San Francisco)	Alameda CA, Contra Costa CA, Marin CA, San Francisco CA, San Mateo CA	Alameda CA, Contra Costa CA, Marin CA, San Francisco CA, San Mateo CA

MSA	Represented counties	Counties in our indices
New York City Area (New York)	Fairfield CT, New Haven CT, Bergen NJ, Essex NJ, Hudson NJ, Hunterdon NJ, Mercer NJ, Middlesex NJ, Monmouth NJ, Morris NJ, Ocean NJ, Passaic NJ, Somerset NJ, Sussex NJ, Union NJ, Warren NJ, Bronx NY, Dutchess NY, Kings NY, Nassau NY, New York NY, Orange NY, Putnam NY, Queens NY, Richmond NY, Rockland NY, Suffolk NY, Westchester NY, Pike PA	Fairfield CT, New Haven CT, Bergen NJ, Essex NJ, Hudson NJ, Hunterdon NJ, Mercer NJ, Middlesex NJ, Monmouth NJ, Morris NJ, Ocean NJ, Passaic NJ, Somerset NJ, Sussex NJ, Union NJ, Warren NJ, Bronx NY, Dutchess NY, Kings NY, Nassau NY, New York NY, Orange NY, Putnam NY, Queens NY, Richmond NY, Rockland NY, Suffolk NY, Westchester NY
Washington-Arlington-Alexandria, DC-VA-MD-WV Metropolitan Statistical Area (Washington, D.C.)	District of Columbia DC, Calvert MD, Charles MD, Frederick MD, Montgomery MD, Prince Georges MD, Alexandria City VA, Arlington VA, Clarke VA, Fairfax VA, Fairfax City VA, Falls Church City VA, Fauquier VA, Fredericksburg City VA, Loudoun VA, Manassas City VA, Manassas Park City VA, Prince William VA, Spotsylvania VA, Stafford VA, Warren VA, Jefferson WV	District of Columbia DC, Calvert MD, Charles MD, Frederick MD, Montgomery MD, Prince Georges MD, Alexandria City VA, Arlington VA, Fairfax VA, Falls Church City VA, Fauquier VA, Fredericksburg City VA, Loudoun VA, Manassas City VA, Manassas Park City VA, Prince William VA, Spotsylvania VA, Stafford VA

Note: The table reports the counties included in the ten Metropolitan Statistical Areas (MSAs) underlying the S&P/Case-Shiller indexes. The name of each MSA is abbreviated by that of its major city or county, as indicated in parenthesis.

Table A.2: Data summary

Los Angeles	Boston	Chicago	Denver	Miami	Las Vegas	San Diego	San Francisco	New York	D.C.
Panel A: Data availability									
Full sample start date									
04/01/88	02/01/87	01/08/96	02/01/98	02/01/97	04/01/88	04/01/88	04/01/88	02/01/87	01/10/96
Daily index start date									
03/01/95	03/01/95	09/01/99	05/03/99	04/01/98	01/03/95	01/02/96	01/03/95	01/03/95	06/01/01
Full sample end date									
10/23/12	10/11/12	10/12/12	10/17/12	10/15/12	10/17/12	10/23/12	10/18/12	10/23/12	10/23/12
Panel B: Transactions									
Total transactions									
10,285,770	2,121,471	3,948,706	1,672,669	3,689,159	2,236,138	2,845,804	3,778,446	5,943,114	2,168,018
Single family residential housing transactions									
5,970,536	1,141,930	1,886,433	1,000,785	1,366,745	1,479,872	1,584,732	2,331,860	2,951,031	1,055,537
Arms-length transaction									
2,562,884	975,964	1,157,215	672,512	935,985	915,408	755,440	1,031,261	2,307,079	759,752
After excluding transaction value ¡= 5000 or ¿= 100,000,000									
2,555,165	917,039	1,156,042	671,605	935,178	913,682	754,106	1,030,384	2,271,467	757,675
After excluding houses sold only once									
1,980,740	638,577	659,732	475,481	668,552	729,365	579,152	757,379	1,234,074	459,842

Los Angeles	Boston	Chicago	Denver	Miami	Las Vegas	San Diego	San Francisco	New York	D.C.
Panel B: Transactions (continued)									
After excluding transactions happen within 6 months									
1,627,149	532,761	561,945	374,045	576,810	645,869	510,450	688,869	1,026,836	325,251
After excluding >= 2× standard deviations and >= 6× median transaction values									
1,578,869	514,356	543,038	360,944	561,805	628,790	494,894	665,537	999,284	313,777
Panel C: Sale pairs									
Total pairs									
939,476	294,101	292,737	198,608	321,358	378,093	296,985	397,229	544,326	162,693
After excluding renovation/reconstruction between two sales									
899,573	286,760	292,737	187,977	287,790	226,701	244,059	350,500	540,235	151,203
After excluding abnormal annual returns (less than -50% or more than 100%)									
878,017	272,858	277,160	181,633	281,393	221,877	239,232	341,878	512,251	143,481
After excluding sale pairs with second transaction on weekends									
878,002	272,727	277,095	180,504	277,442	221,876	239,215	341,858	508,860	143,433
After excluding sale pairs with second transaction on federal holidays									
877,885	272,414	277,079	180,003	276,676	221,554	239,041	341,469	508,548	143,431
Average *daily* sale pairs for the daily index estimation period									
180	55	84	53	77	49	51	70	109	49

Table A.3: Noise filter estimates

	Los Angeles	Boston	Chicago	Denver	Miami	Las Vegas	San Diego	San Francisco	New York	D.C.
μ	0.018	0.019	0.002	0.009	0.013	0.001	0.020	0.018	0.016	0.017
	(0.006)	(0.007)	(0.001)	(0.008)	(0.010)	(0.000)	(0.006)	(0.007)	(0.005)	(0.009)
σ_η	2.457	5.888	4.668	3.779	4.113	5.362	3.746	4.925	4.349	4.612
	(0.039)	(0.140)	(0.155)	(0.139)	(0.076)	(0.212)	(0.058)	(0.105)	(0.132)	(0.108)
σ_u	0.379	0.388	0.593	0.327	0.497	0.568	0.407	0.525	0.376	0.501
	(0.022)	(0.034)	(0.057)	(0.040)	(0.035)	(0.056)	(0.022)	(0.029)	(0.034)	(0.037)
σ_η/σ_u	6.478	15.180	7.866	11.544	8.273	9.448	9.204	9.376	11.576	9.200

Note: Quasi Maximum Likelihood Estimates (QMLE) with robust standard errors in parentheses.

Table A.4: Daily HAR-X-GARCH-CCC correlations

	Los Angeles	Boston	Chicago	Denver	Miami	Las Vegas	San Diego	San Francisco	New York	D.C.
Los Angeles	1.000	-0.001	-0.004	-0.009	0.049	0.033	0.056	0.177	0.011	0.031
Boston		1.000	-0.013	0.014	0.023	0.025	0.041	0.023	-0.010	0.022
Chicago			1.000	-0.004	0.018	0.006	-0.014	0.036	0.048	0.015
Denver				1.000	0.030	0.024	0.031	-0.004	-0.022	0.023
Miami					1.000	0.017	0.025	0.038	0.038	0.018
Las Vegas						1.000	0.028	0.030	-0.023	0.026
San Diego							1.000	0.054	0.002	0.011
San Francisco								1.000	0.005	0.023
New York									1.000	0.025
Washington, D.C.										1.000

Note: Conditional daily correlations estimated from the HAR-X-GARCH-CCC model.

Table A.5: Unconditional correlations of monthly S&P/Case-Shiller index returns

	Los Angeles	Boston	Chicago	Denver	Miami	Las Vegas	San Diego	San Francisco	New York	D.C.
Los Angeles	1.000	0.651	0.658	0.543	0.870	0.875	0.926	0.835	0.778	0.881
Boston		1.000	0.767	0.749	0.527	0.495	0.672	0.693	0.725	0.773
Chicago			1.000	0.679	0.637	0.544	0.567	0.688	0.818	0.762
Denver				1.000	0.398	0.382	0.545	0.693	0.496	0.666
Miami					1.000	0.799	0.782	0.743	0.795	0.802
Las Vegas						1.000	0.819	0.663	0.684	0.748
San Diego							1.000	0.833	0.712	0.839
San Francisco								1.000	0.659	0.855
New York									1.000	0.816
Washington, D.C.										1.000

Note: The correlations are based on the same June 2001 to September 2012 sample period used in the estimation of the daily HAR-X-GARCH-CCC model.

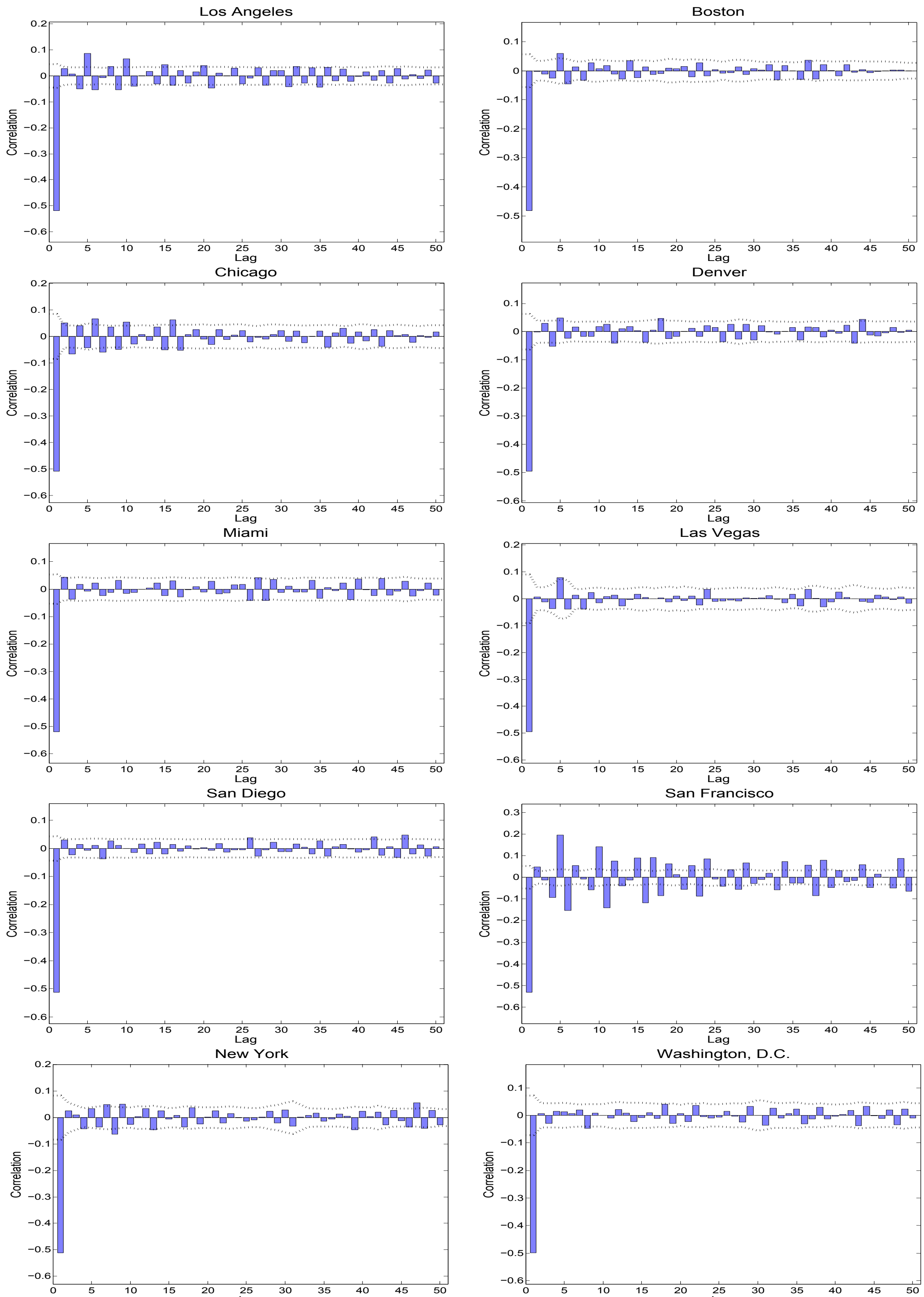


FIGURE A.1: Sample autocorrelations for raw daily index returns

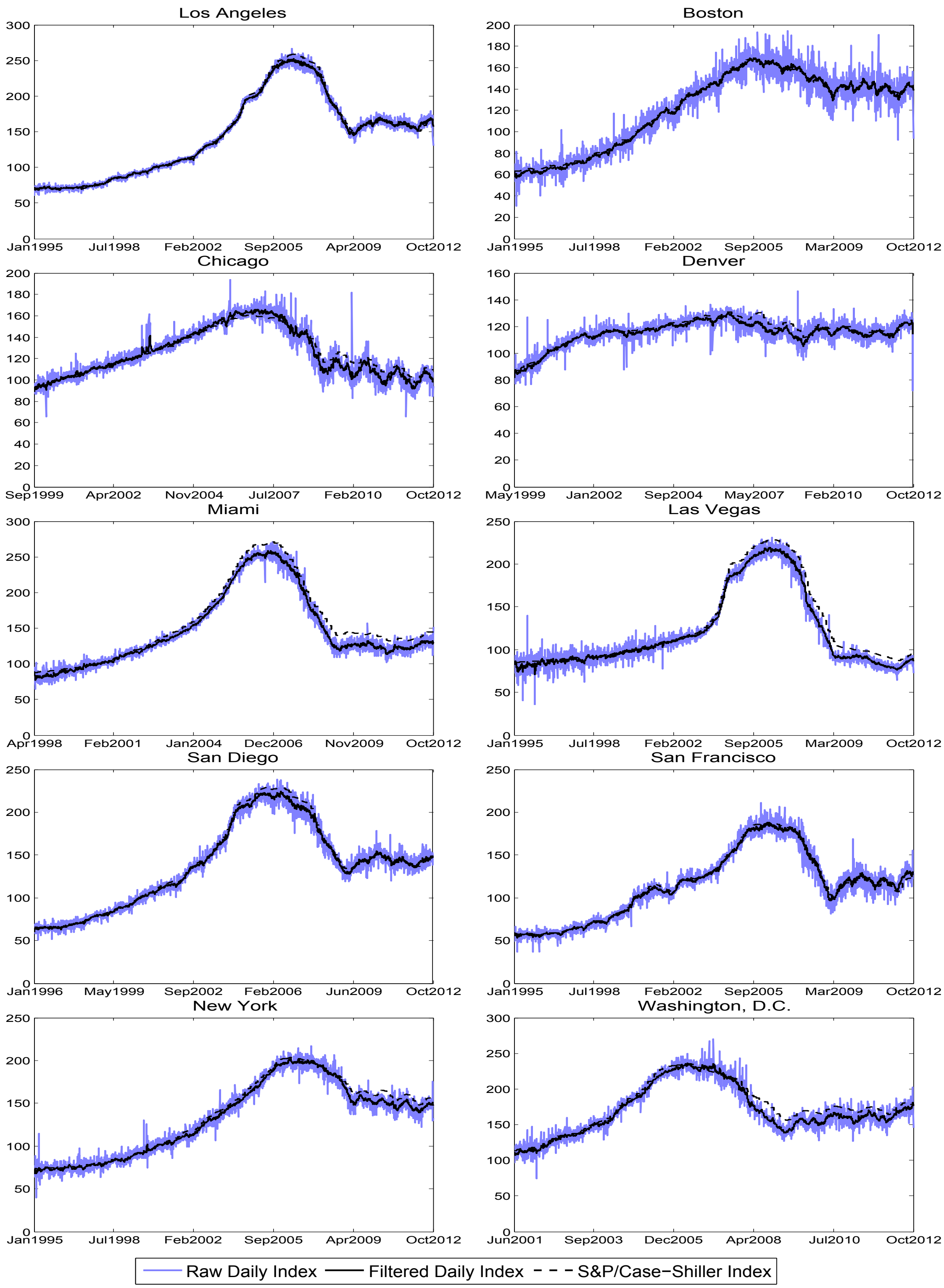


FIGURE A.2: Raw and filtered daily house price indexes for ten MSAs

FIGURE A.3: Unconditional return correlations as a function of return horizon

Appendix B

Appendix to Chapter 4

The stylized equilibrium model discussed in Section 4.4.5 is designed to illustrate the economic mechanism behind the observed volatility asymmetry in aggregate housing and stock volatilities, as well as to shed light on the theoretical linkage between the two volatilities. The model involves a standard endowment economy with Epstein-Zin-Weil recursive preference (Epstein and Zin, 1989; Weil, 1989), which can induce an endogenous volatility risk premium (Bansal and Yaron, 2004; Tauchen, 2011). It builds on the Housing-CAPM model proposed by Piazzesi et al. (2007) and the discrete-time long-run risk model pioneered by Bansal and Yaron (2004). Instead of looking for an exact model solution under a special case, as in Fillat (2008), this paper follows the approximation approach in Bansal and Yaron (2004) to obtain the model solution in a general setting.

Consider an economy with a representative agent that derives utility from a consumption bundle, $\tilde{C}_t$. The recursive utility function is

$$V_t = \left[(1-\delta)\tilde{C}_t^{1-\rho} + \delta E_t[V_{t+1}^{1-\gamma}]^{\frac{1-\rho}{1-\gamma}}\right]^{\frac{1}{1-\rho}} \tag{B.0.1}$$

where δ is the rate of time preference and $\psi \equiv 1/\rho$ is the intertemporal substitution. The parameter γ is the risk aversion coefficient, which determines the curvature of the value function.

Following Piazzesi et al. (2007), the consumption bundle $\tilde{C}_t$ is composed of two goods: consumption of housing services, S_t, and non-housing consumption, C_t.

$$\tilde{C}_t = \left(C_t^{\frac{\varepsilon-1}{\varepsilon}} + \omega S_t^{\frac{\varepsilon-1}{\varepsilon}} \right)^{\frac{\varepsilon}{\varepsilon-1}} \tag{B.0.2}$$

where ε is the elasticity of substitution between housing and non-housing consumption and ω is a preference shift parameter. Housing is incorporated in the model both as an argument of the utility function and as an asset. To simplify the analysis and focus on the volatility dynamics, assume that both the housing and nonhousing consumption can be adjusted without cost[1].

The static first order condition takes the form

$$\frac{P_t^C}{P_t^S} = \frac{1}{\omega} \left(\frac{C_t}{S_t} \right)^{-\frac{1}{\varepsilon}} \tag{B.0.3}$$

where P_t^C and P_t^S are prices for the non-housing consumption and housing service, respectively. Therefore, at the optimal, the consumption bundle can be written as follows:

$$\tilde{C}_t = C_t \left[1 + \omega^{\varepsilon} \left(\frac{P_t^S}{P_t^C} \right)^{1-\varepsilon} \right]^{\frac{\varepsilon}{\varepsilon-1}} \tag{B.0.4}$$

After the approximation and log-linearization, the geometric growth rate of aggregate consumption ($\tilde{g}_{t+1} = \log(\tilde{C}_{t+1}/\tilde{C}_t)$) is the weighted sum of the growth rate of non-housing consumption ($g_{t+1} = \log(C_{t+1}/C_t)$) and the growth rate of relative prices ($g_{p,t+1} = \log(P_{t+1}^S/P_{t+1}^C) - \log(P_t^S/P_t^C)$).

$$\tilde{g}_{t+1} = g_{t+1} + G_1 g_{p,t+1} \tag{B.0.5}$$

where $G_1 = -\omega^{1-\varepsilon}\varepsilon(\overline{\frac{P_t^S}{P_t^C}})^{1-\varepsilon} \Big/ (\overline{\frac{\tilde{C}_t}{C_t}})^{\frac{\varepsilon-1}{\varepsilon}}$ is a linearization constant.

[1] Flavin and Nakagawa (2008) and Flavin (2012) assume that the household incurs an adjustment cost when altering the holding of the durable good (or house), although financial assets and consumption of the nondurable good can be adjusted without cost.

Next, assume the dynamics of the growth rate of non-housing consumption ($g_{t+1} \equiv \log(C_{t+1}/C_t)$), relative price of housing service to nonhousing consumption[2] ($p_t^r \equiv \log(P_t^r) = \log(P_t^S/P_t^C)$) and stock dividend growth rate ($g_{d,t+1} \equiv \log(D_{t+1}/D_t)$) as:

$$g_{t+1} = \sigma_{g,t} z_{g,t+1} \tag{B.0.6}$$

$$\sigma_{g,t+1}^2 = a_\sigma + \rho_\sigma \sigma_{g,t}^2 + \sqrt{q} z_{\sigma,t+1} \tag{B.0.7}$$

$$p_{t+1}^r = \mu_p + \rho_p p_t^r + \sigma_{p,t} z_{p,t+1} \tag{B.0.8}$$

$$\sigma_{p,t+1}^2 = a_{\sigma p} + \rho_{\sigma p} \sigma_{p,t}^2 + \sqrt{q_p} z_{\sigma p,t+1} \tag{B.0.9}$$

$$g_{d,t+1} = k g_{t+1} + \sigma_d z_{d,t+1} \tag{B.0.10}$$

where the parameters satisfy $a_\sigma > 0$, $a_{\sigma p} > 0$, $|\rho_\sigma| < 1$, $|\rho_{\sigma p}| < 1$, $|\rho_p| < 1$, and $z_{g,t+1}$, $z_{\sigma,t+1}$, $z_{p,t+1}$, $z_{\sigma p,t+1}$, $z_{d,t+1}$ are i.i.d. N(0,1) processes. The growth rate of nonhousing consumption g_{t+1} is assumed to be unpredictable with conditional variance $\sigma_{g,t}$ and constant volatility-of-volatility q. The relative price p_t^r follows an autoregressive process with flexible degrees of persistence ρ_p, along with conditional variance $\sigma_{p,t}$ and constant volatility-of-volatility q_p. The specification of dividend growth rate $g_{d,t+1}$ is used for pricing equity later on with $k > 0$ and $\sigma_d \geqslant 0$, to match the empirical variance of stock dividend growth rate and its correlation to nonhousing consumption growth rate[3].

The logarithm of the intertemporal marginal rate of substitution can be expressed as

$$m_{t+1} \equiv \log(M_{t+1}) = \theta \log \delta - \theta \psi^{-1} \tilde{g}_{t+1} + (\theta - 1)\tilde{r}_{t+1} \tag{B.0.11}$$

where $\theta \equiv (1-\gamma)(1-\psi^{-1})^{-1}$ and $\tilde{r}_{t+1}$ is the market return.

Let w_t denote the logarithm of the price-dividend ratio of the asset that pays the consumption bundle endowment, $\{\tilde{C}_{t+j}\}_{j=1}^{\infty}$. Following the method in Bansal and Yaron (2004), we can conjecture a solution for w_t as an affine function of the state variables, $\sigma_{g,t}^2$,

[2] Note that the static first order condition in (B.0.3) implies that specifying the dynamics of equilibrium relative quantity of housing and nonhousing consumption is equivalent to specifying the dynamics of equilibrium relative price.

[3] Piazzesi et al. (2007) provides a detailed discussion on the asset pricing implication according to different specifications of the dividend growth rate.

$\sigma^2_{p,t}$ and p^r_t.

$$w_t = A_0 + A_1\sigma^2_{g,t} + A_2\sigma^2_{p,t} + A_3 p^r_t \tag{B.0.12}$$

The coefficients A_0, A_1, A_2 and A_3 can be solved using Campbell and Shiller (1988) approximation

$$\tilde{r}_{t+1} = \kappa_0 + \kappa_1 w_{t+1} - w_t + \tilde{g}_{t+1} \tag{B.0.13}$$

Since $E_t\left[exp\{m_{t+1} + \tilde{r}_{t+1}\}\right] = 1$, the resulting equilibrium solution for the coefficients are

$$\begin{aligned}
A_1 &= \frac{\theta(\psi^{-1}-1)^2}{2(1-\kappa_1\rho_\sigma)} \\
A_2 &= \frac{\theta\left[(1-\psi^{-1})G_1 + \kappa_1 A_3\right]^2}{2(1-\kappa_1\rho_{\sigma p})} \\
A_3 &= \frac{(\psi^{-1}-1)G_1(\rho_p - 1)}{\kappa_1\rho_p - 1}
\end{aligned} \tag{B.0.14}$$

It then follows that total market return, $\tilde{r}_{t+1}$, takes the following form:

$$\begin{aligned}
\tilde{r}_{t+1} &= \kappa_0 + \kappa_1\left[A_0 + A_1(a_\sigma + \rho_\sigma\sigma^2_{g,t} + \sqrt{q}z_{\sigma,t+1}) + A_2(a_{\sigma p} + \rho_{\sigma p}\sigma^2_{p,t} + \sqrt{q_p}z_{\sigma p,t+1})\right. \\
&\quad \left. + A_3(\mu_p + \rho_p p^r_t + \sigma_{p,t}z_{p,t+1})\right] - (A_0 + A_1\sigma^2_t + A_2\sigma^2_{p,t} + A_3 p^r_t) + \sigma_{g,t}z_{g,t+1} \\
&\quad + G_1\left[\mu_p + (\rho_p - 1)p^r_t + \sigma_{p,t}z_{p,t+1}\right]
\end{aligned} \tag{B.0.15}$$

and the stochastic discount factor (SDF) is

$$\begin{aligned}
m_{t+1} &= \theta\log\delta - \theta\psi^{-1}\tilde{g}_{t+1} + (\theta-1)\tilde{r}_{t+1} \\
&= \theta\log\delta - \theta\psi^{-1}\Big(\mu_g + \sigma_{g,t}z_{g,t+1} + G_1\mu_r + G_1(\rho_r - 1)p_{r,t} + G_1\sigma_{p,t}z_{p,t+1}\Big) \\
&\quad + (\theta-1)\Big(\kappa_0 + \kappa_1\left[A_0 + A_1(a_\sigma + \rho_\sigma\sigma^2_{g,t} + \sqrt{q}z_{\sigma,t+1})\right. \\
&\quad \left. + A_2(a_{\sigma p} + \rho_{\sigma p}\sigma^2_{p,t} + \sqrt{q_p}z_{\sigma p,t+1}) + A_3(\mu_p + \rho_p p^r_t + \sigma_{p,t}z_{p,t+1})\right] \\
&\quad - (A_0 + A_1\sigma^2_t + A_2\sigma^2_{p,t} + A_3 p^r_t) \\
&\quad + \sigma_{g,t}z_{g,t+1} + G_1\left[\mu_p + (\rho_p - 1)p^r_t + \sigma_{p,t}z_{p,t+1}\right]\Big)
\end{aligned} \tag{B.0.16}$$

With aggregate consumption ($\tilde{C}_t$) as the numeraire, the housing asset that pays the housing service S_{t+1} as a dividend at time $t+1$ can be expressed as

$$D^h_{t+1} \equiv \frac{P^S_{t+1} S_{t+1}}{P_{t+1}} \tag{B.0.17}$$

where the price index P_t of the aggregate consumption is

$$P_t = ((P^C_t)^{1-\varepsilon} + \omega^\varepsilon (P^S_t)^{1-\varepsilon})^{1/(1-\varepsilon)} \tag{B.0.18}$$

Next, the logarithm of dividend growth rate of the housing asset $g_{h,t+1}$ can be approximated as

$$g_{h,t+1} \equiv \log(D^h_{t+1}) - \log(D^h_t) = g_{t+1} + (1-\varepsilon-\alpha) g_{p,t+1} \tag{B.0.19}$$

where $\alpha = \omega^\varepsilon \overline{\frac{P^S_t}{P^C_t}}^{1-\varepsilon} \Big/ \overline{\frac{P_t}{P^C_t}}^{1-\varepsilon}$. Then, the Campbell and Shiller (1988) approximation of the housing asset return is

$$r_{h,t+1} = \kappa_{h,0} + \kappa_{h,1} w_{h,t+1} - w_{h,t} + g_{h,t+1} \tag{B.0.20}$$

where $w_{h,t}$ is the logarithmic price-dividend ratio of the asset that pays the housing consumption. Similar to the previous calculation, if we conjecture $w_{h,t} = B_0 + B_1 \sigma^2_{g,t} + B_2 \sigma^2_{p,t} + B_3 p^r_t$ and use $E_t\left[exp\{m_{t+1} + r_{h,t+1}\}\right] = 1$, the equilibrium solutions for the coefficients are as follows:

$$\begin{aligned}
B_2 &= -\frac{G_1^2\theta^2 + 2G_1\psi\theta\left[-1+\alpha+\varepsilon+G_1+A_3\kappa_1 - B_3\kappa_{h,1} - (G_1+A_3\kappa_1)\theta\right]}{2\psi^2(\kappa_{h,1}\rho_{\sigma p}-1)} \\
&- \frac{2A_2(\kappa_1\rho_{\sigma p}-1)(\theta-1) + \left[-1+\alpha+\varepsilon+G_1+A_3\kappa_1 - B_3\kappa_{h,1} - (G_1+A_3\kappa_1)\theta\right]^2}{2(\kappa_{h,1}\rho_{\sigma p}-1)} \\
B_3 &= \frac{G_1\psi^{-1}(\rho_p-1)\theta - A_3(\kappa_1\rho_p-1)(\theta-1) - (\rho_p-1)(-1+\alpha+\varepsilon+G_1-G_1\theta)}{\kappa_{h,1}\rho_p-1} \\
B_1 &= \frac{2A_1(\kappa_1\rho_\sigma-1)(\theta-1) + (1-\psi^{-1})^2\theta^2}{2(1-\kappa_{h,1}\rho_\sigma)}
\end{aligned} \tag{B.0.21}$$

To price equity, assume the stock dividend growth rate as $\log(D_{t+1}/D_t) = k\log(C_{t+1}/C_t) + \sigma_d z_{d,t+1}$. Similar to the pricing of the housing asset, the stock dividend in numeraire of consumption bundle is denoted as D^s_{t+1}.

$$D^s_{t+1} \equiv \frac{P^C_{t+1} D_{t+1}}{P_{t+1}} \tag{B.0.22}$$

Then, the stock dividend growth rate $g_{s,t+1} = \log(D^s_{t+1}/D^s_t)$ can be expressed as

$$g_{s,t+1} = kg_{t+1} + \sigma_d z_{d,t+1} + (-\alpha) g_{p,t+1} \tag{B.0.23}$$

Similarly, approximate the stock returns as $r_{s,t+1} = \kappa_{s,0} + \kappa_{h,1} w_{s,t+1} - w_{s,t} + g_{s,t+1}$ and conjecture $w_{s,t} = C_0 + C_1\sigma^2_{g,t} + C_2\sigma^2_{p,t} + C_3 p^r_t$. Then, solve for the equilibrium solutions of the coefficients.

$$\begin{aligned}
C_1 &= -\frac{\theta^2\psi^{-2} - 2\psi^{-1}\theta(-1+k+\theta) + 2A_1(-1+\kappa_1\rho_\sigma)(-1+\theta) + (-1+k+\theta)^2}{2(-1+\kappa_{s,1}\rho_\sigma)} \\
C_2 &= -\frac{2A_2(\kappa_1\rho_{\sigma p} - 1)(\theta - 1) + \left[\alpha + G_1 + A_3\kappa_1 - C_3\kappa_{s,1} + (G_1\psi^{-1} - G_1 + A_3\kappa_1)\theta\right]^2}{2(-1+\kappa_{s,1}\rho_{\sigma p})} \\
C_3 &= \frac{-A_3\psi(\kappa_1\rho_p - 1)(\theta - 1) + (-1+\rho_p)\left[\alpha\psi + G_1(\psi + \theta - \psi\theta)\right]}{\psi(-1+\kappa_{s,1}\rho_p)}
\end{aligned} \tag{B.0.24}$$

Following much of the stochastic volatility literature, the capital returns of stock and housing, instead of total returns with dividends, are examined:

$$\begin{aligned}
\Delta p_{s,t+1} &\equiv \log(P_{s,t+1}) - \log(P_{s,t}) \\
&= w_{s,t+1} - w_{s,t} + g_{s,t+1} \\
&= (a_\sigma C_1 + a_{\sigma p} C_2 - \alpha\mu_p + C_3\mu_p) - (\alpha - C_3)(-1+\rho_p)p^r_t \\
&\quad + C_1(-1+\rho_\sigma)\sigma^2_{g,t} + C_2(-1+\rho_{\sigma p})\sigma^2_{p,t} + k\sigma_{g,t} z_{g,t+1} \\
&\quad + (-\alpha + C_3)\sigma_{p,t} z_{p,t+1} + \sigma_d z_{d,t+1} \\
&\quad + C_2\sqrt{q_p} z_{\sigma p,t+1} + C_1\sqrt{q} z_{\sigma,t+1}
\end{aligned} \tag{B.0.25}$$

$$
\begin{aligned}
\Delta p_{h,t+1} &\equiv \log(P_{h,t+1}) - \log(P_{h,t}) \\
&= w_{s,t+1} - w_{s,t} + g_{s,t+1} \\
&= (a_\sigma B_1 + a_{\sigma p} B_2 + \mu_p - \alpha\mu_p + B_3\mu_p - \varepsilon\mu_p) \\
&\quad -(-1+\alpha - B_3 + \varepsilon)(-1+\rho_p)p_t^r + B_1(-1+\rho_\sigma)\sigma_{g,t}^2 \\
&\quad +B_2(-1+\rho_{\sigma p})\sigma_{p,t}^2 + \sigma_{g,t} z_{g,t+1} + (1-\alpha+B_3-\varepsilon)\sigma_{p,t} z_{p,t+1} \\
&\quad +B_2\sqrt{q_p} z_{\sigma p,t+1} + B_1\sqrt{q} z_{\sigma,t+1} \qquad \text{(B.0.26)}
\end{aligned}
$$

where $P_{s,t}$ and $P_{h,t}$ are prices of stock and housing asset at time t. The one-step conditional variance processes of the capital returns are defined as

$$
\begin{aligned}
\sigma_{s,t}^2 &\equiv Var_t(\Delta p_{s,t+1}) = C_1^2 q + C_2^2 q_p + \sigma_d^2 + k^2\sigma_{g,t}^2 + (\alpha - C_3)^2\sigma_{p,t}^2 \qquad \text{(B.0.27)} \\
\sigma_{h,t}^2 &\equiv Var_t(\Delta p_{h,t+1}) = B_1^2 q + B_2^2 q_p + \sigma_{g,t}^2 + (-1+\alpha-B_3+\varepsilon)^2\sigma_{p,t}^2 \qquad \text{(B.0.28)}
\end{aligned}
$$

Both the conditional variances $\sigma_{s,t}^2$ and $\sigma_{h,t}^2$ depend on the conditional volatility of nonhousing consumption growth rates, conditional volatility of relative prices, and their volatility-of-volatilities q and q_p.

Volatility asymmetry typically refers to the correlation between the conditional variance process and lagged return process. The model implies the following conditional moments for each market:

$$
\begin{aligned}
Cov_t(\Delta p_{s,t+1}, \sigma_{s,t+1}^2) &= C_1 k^2 q + C_2(\alpha - C_3)^2 q_p \qquad \text{(B.0.29)} \\
Cov_t(\Delta p_{h,t+1}, \sigma_{h,t+1}^2) &= B_1 q + B_2(-1+\alpha-B_3+\varepsilon)^2 q_p \qquad \text{(B.0.30)}
\end{aligned}
$$

The sign of the conditional covariances are not obvious, because it depends on the sign and/or the magnitude of coefficient constants ($B_{j=1,2,3}$, $C_{j=1,2,3}$) and two volatility-of-volatilities (q, q_p). To examine the possible sign of the conditional covariances, the values of coefficient constants are calculated based on model parameter values that have been commonly used or previously estimated in the literature, together with parameter values implied by the data from the National Income and Product Account (NIPA) tables[4]. Under the benchmark parameter values shown in Table B.1, $Cov_t(\Delta p_{h,t+1}, \sigma_{h,t+1}^2)$

[4] The quantities and prices of housing service and nonhousing consumption are quarterly indexes

is always negative, but the sign of $Cov_t(\Delta p_{s,t+1}, \sigma^2_{s,t+1})$ depends on the magnitude of volatility-of-volatility ratio q/q_p. In particular, if q/q_p exceeds the threshold of 4.395×10^{-4}, $Cov_t(\Delta p_{s,t+1}, \sigma^2_{s,t+1})$ is negative.

$$Cov_t(\Delta p_{s,t+1}, \sigma^2_{s,t+1}) < 0 \quad \text{if } q/q_p > 4.395 \times 10^{-4} \tag{B.0.31}$$

$$Cov_t(\Delta p_{h,t+1}, \sigma^2_{h,t+1}) < 0 \tag{B.0.32}$$

The volatility-of-volatility ratio q/q_p is not easily interpretable, and direct estimation would require the use of latent variable techniques. Instead, as a way to gauge the magnitude of q/q_p, the robust estimates for $\sigma^2_{g,t}$ and $\sigma^2_{p,t}$ are first calculated by exponentially smoothing the squared nonhousing consumption growth rate g_{t+1} and squared residuals from AR(1) model of p^r_{t+1} using a smoothing factor of 0.06. The resulting $\hat{\sigma}^2_{g,t}$ and $\hat{\sigma}^2_{g,t}$ are then fitted in AR(1) models, from which the variance ratio of residuals is an estimate for the volatility-of-volatility ratio q/q_p. Under a benchmark parameter value setting, q/q_p is estimated to be 0.218, which is larger than the threshold value of 4.395×10^{-4}. Therefore, $Cov_t(\Delta p_{s,t+1}, \sigma^2_{s,t+1})$ and $Cov_t(\Delta p_{h,t+1}, \sigma^2_{h,t+1})$ should both be negative, which indicates the existence of volatility asymmetries in both markets.

Table B.2 presents the conditions of the volatility-of-volatility ratio for negative covariances, $Cov_t(\Delta p_{s,t+1}, \sigma^2_{s,t+1})$ and $Cov_t(\Delta p_{h,t+1}, \sigma^2_{h,t+1})$ if varying the parameters values of ω and k. ω measures the representative agent's preference towards housing consumption. $\omega = 1$ indicates that the agent regards housing consumption as important as nonhousing consumption, and $\omega > 1$ implies that the agent prefers housing consumption to nonhousing consumption. The empirical stock dividend growth rate is several times higher than the nonhousing consumption growth rate, so k (together with σ_d in (B.0.10)) can be set to match the empirical variance of stock dividend growth rate and its correlation to the nonhousing consumption growth rate. Then, let ω vary around 1 and set k to be 1 or 3

from 1957 to 2012, obtained from NIPA Personal Income and Outlays, Tables 2.3.3 and 2.3.4, lines 8 and 14.

and evaluate the conditions of the volatility-of-volatility ratio q/q_p for negative conditional covariances $Cov_t(\Delta p_{s,t+1}, \sigma^2_{s,t+1})$ and $Cov_t(\Delta p_{h,t+1}, \sigma^2_{h,t+1})$. As shown in Table B.2, for all ω and k examined, the volatility asymmetry should always exist in the housing market. If the agent weights housing consumption more than nonhousing ($\omega > 1$), there would always be an effect of volatility asymmetry in stock market. When the agent puts less or equal weights to housing consumption ($\omega \leqslant 1$), the threshold values of q/q_p for negative $Cov_t(\Delta p_{s,t+1}, \sigma^2_{s,t+1})$ are all very tiny, especially for $k = 3$. The estimated volatility-of-volatility ratio of 0.218 exceeds all the thresholds, which suggests that under regular conditions, $Cov_t(\Delta p_{s,t+1}, \sigma^2_{s,t+1})$ and $Cov_t(\Delta p_{h,t+1}, \sigma^2_{h,t+1})$ should both be negative. This result indicates that volatility asymmetries should be observed in both housing and stock markets.

Next, consider the dynamic relationship of two conditional volatility processes $\sigma^2_{s,t}$ and $\sigma^2_{h,t}$. It follows from (B.0.27) and (B.0.28) that

$$\begin{aligned} Cov_t(\sigma^2_{s,t+1}, \sigma^2_{h,t+1+j}) &= Cov_t(\sigma^2_{s,t+1+j}, \sigma^2_{h,t+1}) \\ &= k^2 \rho^j_\sigma q + (\alpha - C_3)^2(-1 + \alpha - B_3 + \varepsilon)^2 \rho^j_{\sigma p} q_p \\ &> 0 \quad \text{for} \quad j = 0, 1, 2, 3, ... \end{aligned} \tag{B.0.33}$$

The dynamic conditional covariances of volatilities are positive provided that the volatility-of-volatility (q) in the nonhousing consumption growth rate and the volatility-of-volatility (q_p) in the relative price are non-zeros. The covariance peaks at contemporaneity and decays as the length of the lead or lag j increases, although this shape of the covariance as a function of lead or lag j is less likely to be observed empirically, because of the timing mismatch in the data of housing and stock markets, which is discussed in detail in Section 4.5.

Table B.1: Benchmark parameter setting

	Parameter	Source
γ	10	Bansal and Yaron (2004)
ψ	1.5	Bansal and Yaron (2004)
ε	1.27	Piazzesi et al. (2007)
ω	1.039	Flavin and Nakagawa (2008)
ρ_σ	0.958	Data
$\rho_{\sigma p}$	0.976	Data
ρ_p	0.997	Data
$\overline{\frac{P_t^S}{P_t^C}}$	0.779	Data
k	1	Piazzesi et al. (2007)
κ_1	0.997	
$\kappa_{s,1}$	0.997	Bansal and Yaron (2004)
$\kappa_{h,1}$	0.997	

Note: The assumption that $\psi > 1$ is a matter of some debate (see, for example, the discussion in Bansal and Yaron (2004) and its role in leverage effect of equity volatility in Tauchen (2011). The value of $\overline{\frac{P_t^S}{P_t^C}}$ is set to the mean of relative prices from the year of 1959 to 2012. κ_1, $\kappa_{s,1}$ and $\kappa_{h,1}$ are all log-linearization constants, which should be very close to 1.

Table B.2: Conditions for negative covariances of return and variance for stock and housing

	$\omega = 0.8$		$\omega = 1$		$\omega = 1.2$	
	$k = 1$	$k = 3$	$k = 1$	$k = 3$	$k = 1$	$k = 3$
Stock	$q/q_p > 0.099$	$q/q_p > 9.442 \times 10^{-4}$	$q/q_p > 0.011$	$q/q_p > 1.007 \times 10^{-4}$	Always	Always
Housing	Always	Always	Always	Always	Always	Always

Note: This table shows the conditions of volatility-of-volatility ratios q/q_p, for leverage effect in stock asset ($Cov_t(\Delta p_{s,t+1}, \sigma^2_{s,t+1}) < 0$) and in housing asset ($Cov_t(\Delta p_{h,t+1}, \sigma^2_{h,t+1}) < 0$) under parameter values shown in Table B.1 but varying values of ω and k.

Bibliography

Andersen, T. G., Bollerslev, T., Diebold, F. X., and Ebens, H. (2001), "The distribution of realized stock return volatility," *Journal of Financial Economics*, 61, 43–76.

Andersen, T. G., Bollerslev, T., and Diebold, F. X. (2007), "Roughing it up: Including jump components in the measurement, modeling, and forecasting of return volatility," *The Review of Economics and Statistics*, 89, 701–720.

Andersen, T. G., Bollerslev, T., and Meddahi, N. (2011a), "Realized volatility forecasting and market microstructure noise," *Journal of Econometrics*, 160, 220–234.

Andersen, T. G., Bollerslev, T., and Huang, X. (2011b), "A reduced form framework for modeling volatility of speculative prices based on realized variation measures," *Journal of Econometrics*, 160, 176–189.

Bailey, M. J., Muth, R. F., and Nourse, H. O. (1963), "A Regression Method for Real Estate Price Index Construction," *Journal of the American Statistical Association*, 58, 933–942.

Baillie, R. T., Bollerslev, T., and Redfearn, M. R. (1993), "Bear squeezes, volatility spillovers and speculative attacks in the hyperinflation 1920s foreign exchange," *Journal of International Money and Finance*, 12, 511–521.

Bansal, R. and Yaron, A. (2004), "Risks for the long run: A potential resolution of asset pricing puzzles," *The Journal of Finance*, 59, 1481–1509.

Bardhan, A., Karapandža, R., and Urošević, B. (2006), "Valuing mortgage insurance contracts in emerging market economies," *The Journal of Real Estate Finance and Economics*, 32, 9–20.

Barndorff-Nielsen, O. E. (2002), "Econometric analysis of realized volatility and its use in estimating stochastic volatility models," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 64, 253–280.

Black, F. (1976), "Studies of stock price volatility changes," in *Proceedings of the 1976 Meetings of the American Statistical Association, Business and Economics Statistics Section*, pp. 177–181.

Bollerslev, T. (1986), "Generalized autoregressive conditional heteroskedasticity," *Journal of Econometrics*, 31, 307–327.

Bollerslev, T. (1990), "Modelling the Coherence in Short-Run Nominal Exchange Rates: A Multivariate Generalized ARCH Model," *Review of Economics and Statistics*, 72, 498–505.

Bollerslev, T. and Wooldridge, J. M. (1992), "Quasi-maximum likelihood estimation and inference in dynamic models with time-varying covariances," *Econometric reviews*, 11, 143–172.

Bollerslev, T., Litvinova, J., and Tauchen, G. (2006), "Leverage and volatility feedback effects in high-frequency data," *Journal of Financial Econometrics*, 4, 353–384.

Bollerslev, T., Kretschmer, U., Pigorsch, C., and Tauchen, G. (2009), "A discrete-time model for daily S & P500 returns and realized variations: Jumps and leverage effects," *Journal of Econometrics*, 150, 151–166.

Bollerslev, T., Sizova, N., and Tauchen, G. (2012), "Volatility in equilibrium: Asymmetries and dynamic dependencies," *Review of Finance*, 16, 31–80.

Bollerslev, T., Patton, A., and Wang, W. (2013), "Daily House Price Indexes: Construction, Modeling, and Longer-Run Predictions," *Working Paper*.

Calhoun, C. A. (1996), "OFHEO House Price Indexes : HPI Technical Description," *Office of Federal Housing Enterprise Oversight*.

Calhoun, C. A., Chinloy, P., and Megbolugbe, I. F. (1995), "Temporal aggregation and house price index construction," *Journal of Housing Research*, 6, 419–438.

Campbell, J. Y. and Hentschel, L. (1992), "No news is good news: An asymmetric model of changing volatility in stock returns," *Journal of financial Economics*, 31, 281–318.

Campbell, J. Y. and Shiller, R. J. (1988), "The dividend-price ratio and expectations of future dividends and discount factors," *Review of financial studies*, 1, 195–228.

Campbell, J. Y., Lettau, M., Malkiel, B. G., and Xu, Y. (2001), "Have individual stocks become more volatile? An empirical exploration of idiosyncratic risk," *The Journal of Finance*, 56, 1–43.

Caplin, A., Chopra, S., Leahy, J., Lecun, Y., and Thampy, T. (2008), "Machine Learning and the Spatial Structure of House Prices and Housing Returns," Working paper, New York University.

Case, B. and Quigley, J. M. (1991), "The Dynamics of Real Estate Prices," *Review of Economics and Statistics*, 73, 50–58.

Case, K. E. and Shiller, R. J. (1987), "Prices of single-family homes since 1970: New indexes for four cities," *New England Economic Review*, pp. 45–56.

Case, K. E. and Shiller, R. J. (1989), "The Efficiency of the Market for Single-Family Homes," *American Economic Review*, 79, 125–137.

Case, K. E., Quigley, J. M., and Shiller, R. J. (2005), "Comparing Wealth Effects: The Stock Market versus the Housing Market," *Advances in Macroeconomics*, 5, 1–32.

Case, K. E., Quigley, J. M., and Shiller, R. J. (2011), "Wealth Effects Revisited 1978-2009," Working Paper 16848, National Bureau of Economic Research.

Cho, M. (1996), "House price dynamics: A survey of theoretical and empirical issues," *Journal of Housing Research*, 7, 145–172.

Christie, A. A. (1982), "The stochastic behavior of common stock variances: Value, leverage and interest rate effects," *Journal of financial Economics*, 10, 407–432.

Clapp, J. M. and Giaccotto, C. (1992), "Estimating Price Indices for Residential Property: A Comparison of Repeat Sales and Assesses Value Methods," *Journal of the American Statistical Association*, 87, 300–306.

Clayton, J. and MacKinnon, G. (2001), "The time-varying nature of the link between REIT, real estate and financial asset returns," *Journal of Real Estate Portfolio Management*, 7, 43–54.

Cocco, J. F. (2005), "Portfolio choice in the presence of housing," *Review of Financial studies*, 18, 535–567.

Connolly, R. and Stivers, C. (2006), "Information content and other characteristics of the daily cross-sectional dispersion in stock returns," *Journal of Empirical Finance*, 13, 79–112.

Corsi, F. (2009), "A simple approximate long-memory model of realized volatility," *Journal of Financial Econometrics*, 7, 174–196.

Cotter, J. and Roll, R. (2011), "A comparative anatomy of REITs and residential real estate indexes: Returns, risks and distributional characteristics," .

El Hedi Arouri, M., Jouini, J., and Nguyen, D. K. (2011), "Volatility spillovers between oil prices and stock sector returns: Implications for portfolio management," *Journal of International Money and Finance*, 30, 1387–1405.

Engle, R. F. (2002), “Dynamic Conditional Correlation: A Simple Class of Multivariate Generalized Autoregressive Conditional Heteroskedasticity Models,” *Journal of Business & Economic Statistics*, 20, 339–350.

Engle, R. F. and Ng, V. K. (1993), “Measuring and testing the impact of news on volatility,” *The Journal of Finance*, 48, 1749–1778.

Epstein, L. G. and Zin, S. E. (1989), “Substitution, risk aversion, and the temporal behavior of consumption and asset returns: A theoretical framework,” *Econometrica*, pp. 937–969.

Figlewski, S. and Wang, X. (2000), “Is the “Leverage Effect” a Leverage Effect?” *Available at SSRN 256109*.

Fillat, J. (2008), “Housing as a measure for long-run risk in asset pricing,” in *AFA 2009 San Francisco Meetings Paper*.

Flavin, M. and Nakagawa, S. (2008), “A model of housing in the presence of adjustment costs: A structural interpretation of habit persistence,” *The American Economic Review*, 98, 474–495.

Flavin, M. and Yamashita, T. (2002), “Owner-occupied housing and the composition of the household portfolio,” *The American Economic Review*, 92, 345–362.

Flavin, M. and Yamashita, T. (2011), “Owner-occupied housing: Life-cycle implications for the household portfolio,” *The American Economic Review*, 101, 609–614.

Flavin, M. A. (2012), “Housing, Adjustment Costs, and Macro Dynamics,” *CESifo Economic Studies*, 58, 471–494.

Garcia, R., Mantilla-Garcia, D., and Martellini, L. (2011), “Idiosyncratic Risk and the Cross-Section of Stock Returns,” *Working Paper*, 5, 57.

Ghysels, E., Plazzi, A., Torous, W., and Valkanov, R. (2013), “Forecasting Real Estate Prices, in Graham Elliott and Allan Timmermann (eds.),” *Handbook of Economic Forecasting*, p. forthcoming.

Glosten, L. R., Jagannathan, R., and Runkle, D. E. (1993), “On the relation between the expected value and the volatility of the nominal excess return on stocks,” *The journal of finance*, 48, 1779–1801.

Goetzmann, W. N. and Spiegel, M. (1997), “A Spatial Model of Housing Returns and Neighborhood Substitutability,” *Journal of Real Estate Finance and Economics*, 14, 11–31.

Hamao, Y., Masulis, R. W., and Ng, V. (1990), “Correlations in price changes and volatility across international stock markets,” *Review of Financial studies*, 3, 281–307.

Hoesli, M. and Reka, K. (2011), "Volatility spillovers, comovements and contagion in securitized real estate markets," *The Journal of Real Estate Finance and Economics*, pp. 1–35.

Hoesli, M., Lekander, J., and Witkiewicz, W. (2004), "International evidence on real estate as a portfolio diversifier," *Journal of Real Estate Research*, 26, 161–206.

King, M. A. and Wadhwani, S. (1990), "Transmission of volatility between stock markets," *Review of Financial studies*, 3, 5–33.

Lamont, O. and Stein, J. C. (1999), "Leverage and House-Price Dynamics in US Cities," *The Rand Journal of Economics*, pp. 498–514.

Lancaster, K. J. (1966), "A new approach to consumer theory," *The journal of political economy*, pp. 132–157.

Malik, F. and Ewing, B. T. (2009), "Volatility transmission between oil prices and equity sector returns," *International Review of Financial Analysis*, 18, 95–100.

Meese, R. A. and Wallace, N. E. (1997), "The construction of residential housing price indices: A comparison of repeat-sales, hedonic-regression, and hybrid approaches," *Journal of Real Estate Finance and Economics*, 14, 51–73.

Miao, H., Ramchander, S., and Simpson, M. W. (2011), "Return and volatility transmission in US housing markets," *Real Estate Economics*, 39, 701–741.

Michayluk, D., Wilson, P. J., and Zurbruegg, R. (2006), "Asymmetric volatility, correlation and returns dynamics between the US and UK securitized real estate markets," *Real Estate Economics*, 34, 109–131.

Miller, N. and Peng, L. (2006), "Exploring metropolitan housing price volatility," *The Journal of Real Estate Finance and Economics*, 33, 5–18.

Nagaraja, C. H., Brown, L. D., and Zhao, L. H. (2011), "An autoregressive approach to house price modeling," *Annals of Applied Statistics*, 5, 124–149.

Newey, W. K. and West, K. D. (1987), "A Simple, Positive Semi-definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix," *Econometrica*, 55, 703–708.

Newey, W. K. and West, K. D. (1994), "Automatic lag selection in covariance matrix estimation," *The Review of Economic Studies*, 61, 631–653.

Patton, A. J. and Sheppard, K. (2011), "Good volatility, bad volatility: Signed jumps and the persistence of volatility," *Working Paper*.

Pavlov, A. and Wachter, S. (2010), "REITs and Underlying Real Estate Markets: Is There a Link?" *Working Paper*.

Piazzesi, M., Schneider, M., and Tuzel, S. (2007), "Housing, consumption and asset pricing," *Journal of Financial Economics*, 83, 531–569.

Plazzi, A., Torous, W., and Valkanov, R. (2008), "The cross-sectional dispersion of commercial real estate returns and rent growth: Time variation and economic fluctuations," *Real Estate Economics*, 36, 403–439.

Quigley, J. M. (1995), "A Simple Hybrid Model for Estimating Real Estate Price Indexes," *Journal of Housing Economics*, 4, 1–12.

Rappaport, J. (2007), "A guide to aggregate house price measures," *Economic Review, Federal Reserve Bank of Kansas City*, pp. 41–71.

Rosen, S. (1974), "Hedonic prices and implicit markets: product differentiation in pure competition," *The journal of political economy*, pp. 34–55.

Shiller, R. J. (1991), "Arithmetic repeat sales price estimators," *Journal of Housing Economics*, 1, 110–126.

Shiller, R. J. (1993), "Measuring Asset Values for Cash Settlement in Derivative Markets: Hedonic Repeated Measures Indices and Perpetual Futures," *Journal of Finance*, 48, 911–931.

Standard & Poor's (2009), "S&P/Case-Shiller Home Price Indices Methodology," .

Stivers, C. T. (2003), "Firm-level return dispersion and the future volatility of aggregate stock market returns," *Journal of Financial Markets*, 6, 389–411.

Susmel, R. and Engle, R. F. (1994), "Hourly volatility spillovers between international equity markets," *Journal of International Money and Finance*, 13, 3–25.

Tauchen, G. (2011), "Stochastic Volatility in General Equilibrium," *The Quarterly Journal of Finance*, 1, 707–731.

Tauchen, G., Zhang, H., and Liu, M. (1996), "Volume, volatility, and leverage: A dynamic analysis," *Journal of Econometrics*, 74, 177–208.

Tsay, R. S. (2005), *Analysis of Financial Time Series*, John Wiley & Sons.

Van Nieuwerburgh, S. and Weill, P.-O. (2010), "Why has house price dispersion gone up?" *The Review of Economic Studies*, 77, 1567–1606.

Webb, J. R., Curcio, R. J., and Rubens, J. H. (1988), "Diversification gains from including real estate in mixed-asset portfolios," *Decision Sciences*, 19, 434–452.

Weil, P. (1989), "The equity premium puzzle and the risk-free rate puzzle," *Journal of Monetary Economics*, 24, 401–421.

Yao, R. and Zhang, H. H. (2005), “Optimal consumption and portfolio choices with risky housing and borrowing constraints,” *Review of Financial Studies*, 18, 197–239.

Zhu, B., Füss, R., and Rottke, N. B. (2012), “Spatial Linkages in Returns and Volatilities among US Regional Housing Markets,” *Real Estate Economics*.

Biography

Wenjing Wang was born in China, March 29, 1984. She earned her Bachelor's degree in Computer Science and Technology in 2006 from Renmin University of China. She received her Master's degree in Economics in 2008 from Duke University, and Master's degree in Statistics in 2009 from Yale University. Wenjing Wang expects to graduate from Duke University with doctorate degree in Economics in the summer of 2014.

During her graduate studies at Duke University, Wenjing Wang worked as a teaching assistant at the Department of Statistical Science from 2007 to 2008, and served as a teaching and research assistant at the Department of Economics from 2009 to 2013.